DISS. ETH NO._____

Identification and Characterization of Collagen Glycosyltransferases of Human and Viral Origin

A dissertation submitted to
ETH ZURICH

for the degree of
Doctor of Sciences

presented by
BELINDA SERAINA PAMELA SCHEGG

Dipl. Natw. ETH
born on 17.12.1979
citizen of Oberriet-Montlingen SG

accepted on the recommendation of
Prof. Dr. Markus Aebi, examiner
Prof. Dr. Sabine Werner, co-examiner
Prof. Dr. Thierry Hennet, external examiner

2008

Felix, qui potuit rerum cognoscere causas (Vergil)

Happy is who is able to know the causes of things

Table of Contents

Summary

Collagens are a superfamily of glycoproteins mainly found in the extracellular matrix. They are the most abundant proteins in the human body. Collagens are characterized by a right handed triple helix formed out of three left-handed α-chains representing repeats of the motif G-X-Y, where (hydroxy)proline and (hydroxy)lysine are often found at positions X and Y. To act as a functional collagen, selected hydroxylysine residues have to be further modified by the addition of either galactose or the disaccharide glucosylgalactose. This glycosylation of collagen takes place in the endoplasmic reticulum before the formation of the triple helix and is mediated by specific β(1-O) galactosyl- and α(1-2) glucosyltransferase enzymes. The molecular nature of these glycosyltransferases has remained unknown to date. The present study describes the identification of collagen galactosyltransferase enzymes using a strategy based on affinity chromatography and protein sequencing by mass spectrometry. Three structurally related candidate genes were cloned and expressed in Sf9 insect cells using the baculovirus system. Two of the three candidate glycosyltransferases (GLT25D1 and GLT25D2) were confirmed to be active collagen galactosyltransferases. The collagen galactosyltransferase genes are differentially expressed in human tissues, suggesting that these enzymes may show preference for different types of collagens or contribute to the varying extent of collagen glycosylation throughout tissues. This was supported by showing a selective preference of GLT25D1 and GLT25D2 for collagen type III and collagen type IV acceptors. GLT25D1 showed a higher enzymatic activity on deglycosylated collagen type I to type V than GLT25D2.

Collagen glycosylation is conserved in animals and collagen is also found in several prokaryotic genomes. Proteins sharing structural similarity with the collagen galactosyltransferases have been found in prokaryotes and even in virus. The *Acanthamoeba polyphaga* mimivirus has been detected as an unique member of the nucleo-cytoplasmic large DNA virus family being clearly a virus but also showing features never seen before in viruses. Mimivirus is the largest known DNA virus with a 1.2 Mbp linear dsDNA ge-

nome. It was reported that mimivirus encodes eight proteins with a collagen triple helix motif. These collagens are most probably found in the fibrils of the virus capsid. These fibrils cover the whole icosahedral virus capsid, which are specific for mimivirus. As the Gram staining of mimivirus is positive, it is supposed that the viral fibrils are glycosylated. This suggests that the mimiviral collagens might be post-translationally modified by hydroxylation and subsequent glycosylation.

In this study, the protein L230 was identified as a mimiviral collagen glucosyltransferase transferring glucose on the acceptor hydroxylysine on animal and on mimiviral collagens. This addition of glucose on the acceptor hydroxylysine in collagen has not been reported up to now. It seems that collagen glycosylation in mimivirus is different than in animal collagens.

In conclusion, this work reports the identification and characterization of the two human collagen galactosyltransferases GLT25D1 and GLT25D2 and the identification of the mimiviral collagen glucosyltransferase L230.

Zusammenfassung

Kollagene sind Glykoproteine, die hauptsächlich in der extrazellulären Matrix vorkommen und sind daher die am häufigsten vorkommenden Proteine im menschlichen Körper. Kollagene haben eine charakteristische rechts drehende tripel-helikale Struktur. Diese Helix wird aus drei links drehenden α-Ketten aufgebaut, welche aus dem repetitiven Sequenzmotif Gly-X-Y bestehen. Häufig stehen an den Positionen X und Y (Hydroxy)Prolin oder (Hydroxy)Lysin. Damit das Kollagen funktionell aktiv sein kann, muss es mit Galaktose oder Glukosylgalaktose post-translationell modifiziert werden. Diese Kollagen Glykosylierung findet im Endoplasmatischen Retikulum statt bevor die Tripel-Helix gebildet wird und wird von einer spezifischen β(1-O)Galactosyltransferase und einer α(1-2)Glucosyltransferase durchgeführt. Diese beiden Glykosyltransferasen waren bis heute noch nicht bekannt. Diese Arbeit beschreibt die Identifizierung dieser Kollagen Galaktosyltransferasen, welche mit Hilfe einer Affinitätschromatographie und anschliessender Proteinsequenzierung mittels Massenspektroskopie identifiziert werden konnten. Daraufhin wurden drei strukturell verwandte Kandidatengene kloniert und mittels Bakuloviren in Sf9 Insektenzellen exprimiert. Zwei der drei gefundenen Kandidaten (GLT25D1 und GLT25D2) zeigten in einem Aktivitätstest eine Kollagen-Galaktosyltransferase-Aktivität.

Die Kollagen Galaktosyltransferasen sind in menschlichen Geweben unterschiedlich exprimiert. Dies lässt vermuten, dass diese Enzyme eine Präferenz für verschiedene Kollagen Typen zeigen könnten, oder dass sie dazu beitragen, dass die Kollagene in den verschiedenen Geweben verschieden stark glykosyliert werden. Diese Vermutung wurde bestärkt indem gezeigt wurde, dass GLT25D1 und GLT25D2 präferentiell Kollagen Typ III und Kollagen Typ IV glykosylieren. GLT25D1 hingegen zeigte eine höhere enzymatische Aktivität auf deglykosyliertem Kollagen Typ I bis Kollagen Typ V als GLT25D2.

Kollagen Glykosylierung in Tieren ist ein stark konservierter Prozess. Kollagen selbst findet man jedoch nicht nur in höheren Lebewesen, sondern auch in verschiedenen prokaryontischen Organismen. Daher ist es nicht erstaun-

lich, dass auch Proteine mit strukturellen Ähnlichkeiten zu den Kollagen Galaktosyltransferasen in Prokaryonten und sogar in Viren gefunden wurden. Ein Beispiel dafür ist das *Acanthamoeba polyphaga* Mimivirus. Es ist ein einzigartiges Wesen, das auf der einen Seite klar ein Virus ist, jedoch auf der anderen Seite auch charakteristische Merkmale hat, die nie zuvor in Viren, gefunden worden sind. Mit einem 1.2 Mbp grossen linearen doppelsträngigen DNS Genom ist das Mimivirus das grösste bekannte DNS Virus. Es wurde berichtet, dass das Mimivirus acht Proteine kodiert, welche ein Tripel-Helix Kollagen Motiv haben. Diese Kollagene findet man mit grosser Wahrscheinlichkeit in den Fibrillen auf dem Viruskapsid. Die für das Mimivirus spezifischen Fibrillen bedecken das ganze ikosahedrische Viruskapsid. Die Gram-Färbung des Mimivirus ist positiv, was darauf hindeutet, dass die mimiviralen Kollagene durch Hydroxylierung und anschliessende Glykosylierung posttranslationell modifiziert werden.

In dieser Arbeit wurde das Protein L230 als mimivirale Kollagen Glukosyltransferase identifiziert. Wir konnten zeigen, dass L230 die Übertragung von Glukose auf den Hydroxylysin-Akzeptor sowohl in tierischen als auch in mimiviralen Kollagenen katalysiert. Dieser Transfer von Glukose auf den Hydroxylysin-Akzeptor in Kollagen wurde bis jetzt noch nicht beschrieben. Daraus lässt sich schliessen, dass sich die Kollagen Glykosylierung im Mimivirus von der Kollagen Glykosylierung in Tieren unterscheidet.

Zusammenfasst zeigt diese Arbeit die Identifizierung und Charakterisierung der zwei humanen Kollagen Galaktosyltransferasen GLT25D1 und GLT25D2 sowie die Identifizierung der mimiviralen Kollagen Glukosyltransferase L230.

1 Introduction

1.1 Glycosylation

Glycosylation is the most prevalent posttranslational modification. Glycosylation represents the addition of carbohydrates to aglycons like proteins and lipids, thereby leading to the formation of glycoconjugates. These glycoconjugates consist either of monosaccharides, oligosaccharides or polysaccharides, which are attached through a glycosidic bond.

Protein glycosylation is a highly diverse process and gives rise to a large diversity in structural complexity. Even though mammals only use nine distinct monosaccharides for glycan biosynthesis (Figure 1), the diversity in glycosylation exceeds the diversity achieved by other post-translational modifications by far. The multiple glycosidic linkages applicable and the branched nature of glycan chains contribute largely to the diversity of glycoconjugates.

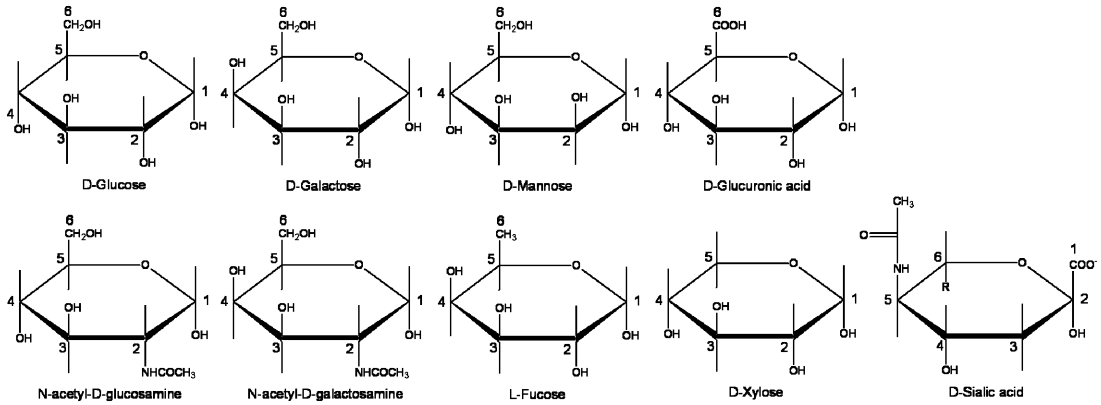


Figure 1: Haworth Projection of the Nine Mammalian Monosaccharides. The chemical structure of the nine human monosaccharides is shown.

1.1.1 Glycan Structure and Conformation

Although the theoretical number of carbohydrate combinations is astronomical, some rules of biosynthesis apply, thereby restricting the repertoire of glycoconjugates to specific classes. The basic monosaccharide units used in mammals are six carbon sugars (hexoses) with the exception of Xylose (Xyl), which has five carbons (pentose) and sialic acid (Sia), which has nine carbon atoms. Four of the six carbon atoms of the hexose are chiral centres, because they have four chemically different substituents. The substituents

around the chiral centre the farthest away from the carbonyl group of the glycan can be arranged in two different stereochemical ways. These stereoisomers are named D- and L-form enantiomers, due to their inability to be converted in each other (mirror images). In contrast, D-glucose (Glc) and D-galactose (Gal) are epimers. They only differ in the stereochemical configuration of the carbon number 4 (C4).

The monosaccharides are linked via a glycosidic bond through C1 and a hydroxyl group of another monosaccharide. C1 has four chemically different substituents and is therefore a chiral centre referred to as anomeric carbon atom. The two possible configurations (anomers) at the anomeric C1 are the α and the β configurations (Figure 2). The structure of a disaccharide is given by its monosaccharide constituents and its linkage. For example lactose is described as Galβ(1-4)Glc where the C1 of Gal is glycosidically linked to the C4 of Glc in a β-linkage.

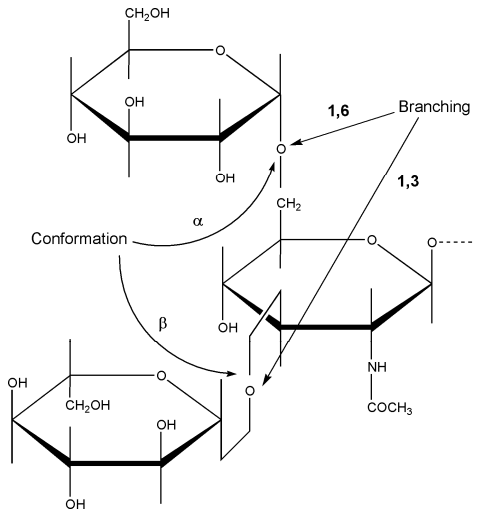


Figure 2: Glycosidic Linkages. The glycosidic linkage can be in an α or in a β conformation. An α(1-6) and a β(1-3) linkage are shown.

1.1.2 Glycoconjugate Classification

Seven classes of glycosylation have been described to date. The main form of protein-bound glycans, the N-glycans, is linked to the amino group of an asparagine residue. O-glycans are linked to the hydroxyl group of serine (S), threonine (T) or hydroxylysine (Hyl) residues (Figure 3). The five additional glycosylation classes represent C-glycans, which are linked through an un-

usual C-C bond to a tryptophane (W) residue (1, 2), the glycosaminoglycans (GAG), which are linked to a serine, the GPI anchor, which consists of phosphatidylinositol (PI) linked to the glycans, the glycolipids, which are linked to ceramide and the cytoplasmic and nuclear glycosylation forms, which represents GlcNAc bound to serine or threonine.

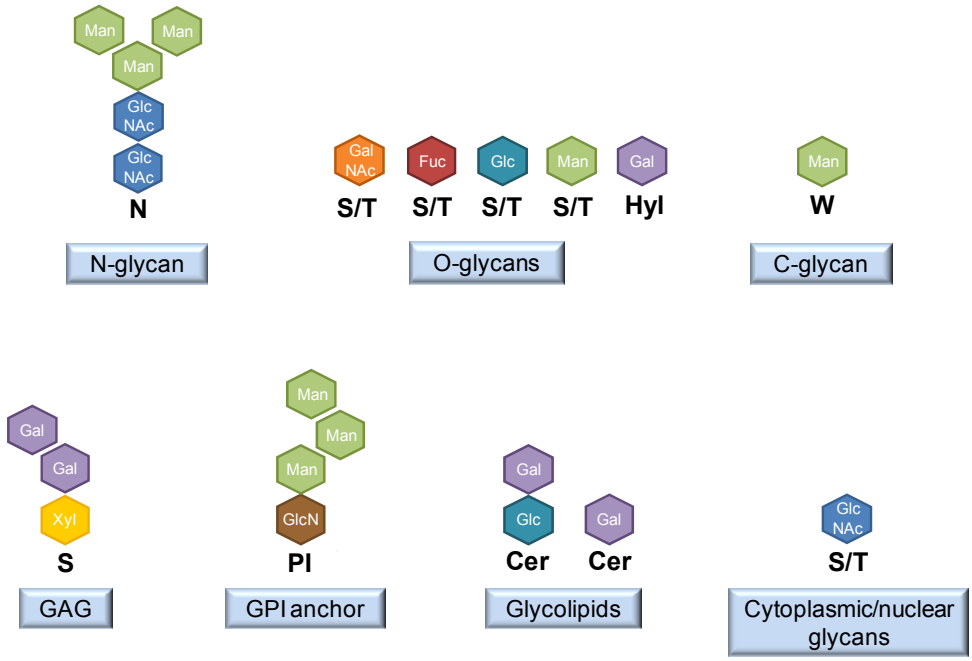


Figure 3: Classification of the Glycoconjugates in Mammals. Shown are the seven classes of glycosylation. N, S, T and W stand for the amino acids asparagine, serine, threonine and tryptophane. PI is phosphatidylinositol and Cer stands for ceramide.

1.1.2.1 N-Glycosylation

Asparagine (N)-linked glycosylation is the best understood glycosylation pathway. The synthesis comprises several steps which can be grouped into three major stages. The process is initiated by the formation of a lipid-linked oligosaccharide (LLO), followed by an *en bloc* transfer to the nascent polypeptide in the sequence Asn-X-Ser/Thr, where X can be every amino acid except proline, and is terminated by a final processing of the oligosaccharide. The elongation of the lipid carrier dolichol takes place at the membrane of the endoplasmic reticulum (ER). The first five monosaccharide units are added at the cytosolic side of the ER membrane. After flipping to the luminal side of the membrane, the LLO assembly proceeds with the sequential addition of nine monosaccharides (Figure 4). The complete structure $Glc_3Man_9GlcNAc_2$

is co-translationally transferred *en bloc* by the oligosaccharyltransferase on an asparagine residue of a nascent polypeptide. After transfer on proteins, glucosidases remove two of the three terminal Glcs. The third Glc is important for the quality control cycle in the ER and is removed during this process (3, 4). Correctly folded proteins carrying $Man_8GlcNAc_2$ glycans are further transported to the Golgi apparatus, where N-glycans are further trimmed down and elongated (for further details on N-glycosylation see review (3)). While the N-glycans leaving the ER always have the same structure, the glycosylation steps in the Golgi result in a tremendous diversity of N-linked glycans.

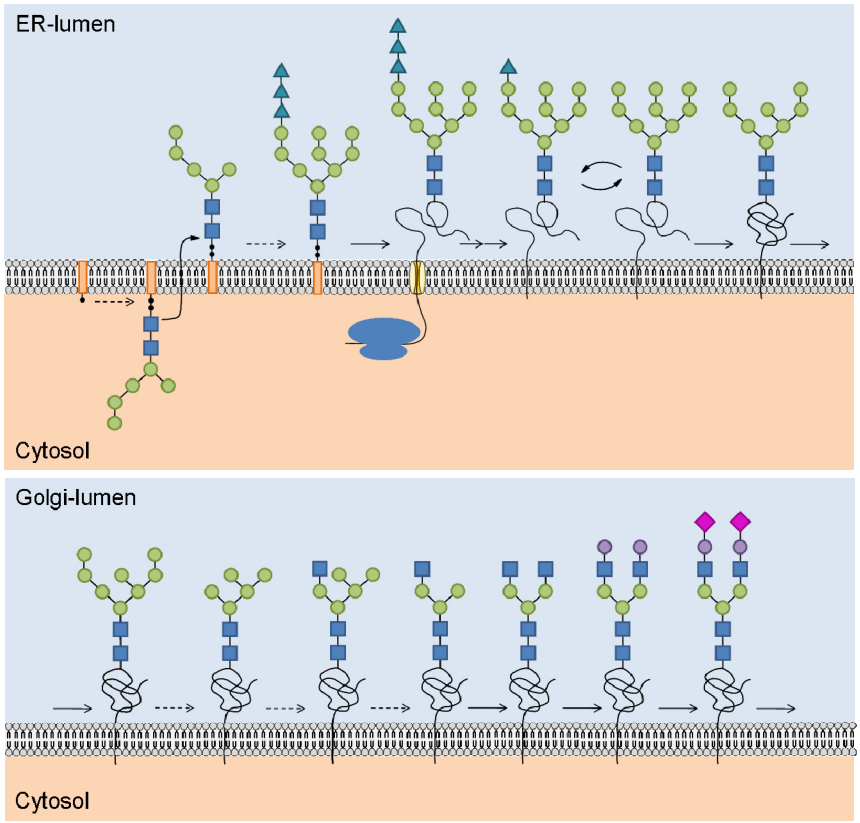


Figure 4: Scheme of N-Glycosylation. N-glycosylation starts on the cytosolic side of the ER with the assembly of the LLO $Man_5GlcNAc_2$ which is then flipped into the lumen of the ER. $Glc_3Man_9GlcNAc_2$ is transferred on the nascent polypeptide followed by the removal of the three terminal Glcs by ER-glucosidases. $Man_8GlcNAc_2$ is transported to the Golgi where further trimming and subsequent modifications of the oligosaccharide take place.

1.1.2.2 O-Glycosylation

In contrast to N-glycosylation where oligosaccharides are transferred *en bloc* to proteins, O-linked glycosylation is initiated by the transfer of single mono-

saccharides. The most common O-glycan is a GalNAc (5-7) linked to the hydroxyl group of either serine or threonine, known as the T_n-antigen, which is then further elongated with other monosaccharides. This type of O-linked glycosylation takes place in the Golgi apparatus (8). The other types of O-linked glycosylation begin in the ER. To these groups belong O-linked Fuc and Glc, added in the ER, which are found in a specific consensus sequence in the epidermal growth factor-like (EGF) and thrombospondin domains (9). Additionally, O-linked mannose on serine or threonine is synthesized in the ER (8, 10). Glycosylation is not confined to the organelles of the secretory pathway. A reversible glycosylation implying O-linked GlcNAc is also found in the cytoplasm and in the nucleus (11). Finally, a type of O-linked glycosylation is specific to the collagen domain. This O-linked addition of Gal or of the disaccharide Glc-Gal is discussed in chapter 1.2.4.2.

1.1.3 Glycosyltransferases

The synthesis of glycans is mediated by glycosyltransferase enzymes (GT). GTs catalyze the formation of glycosidic linkages to either connect two sugars or a sugar with an aglycon. GTs use activated monosaccharides as donor substrates. This activation is achieved by binding to nucleotides or to dolichol-phosphate. Typical nucleotide-activated sugars are UDP-Gal, GDP-Man or CMP-Sia.

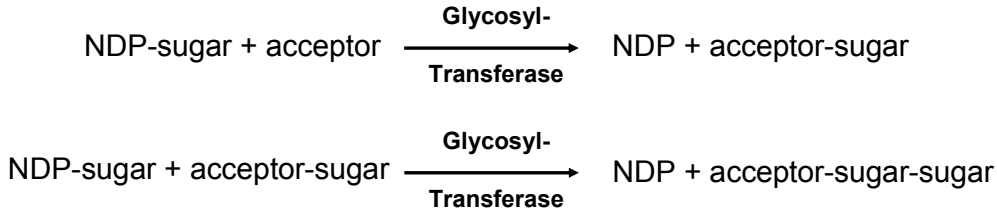


Figure 5: Glycosyltransferase Reaction. Activated nucleotide sugars are hydrolyzed by the glycosyltransferase, catalyzing the formation of a glycosidic linkage and the release of NDP (nucleotide diphosphate) or in the case of CMP-Sia the release of NMP (nucleotide monophosphate).

After synthesis in the cytosol and in the nucleus (CMP-sialic acid), nucleotide-sugars are transported across the membrane into the lumen of the ER or the Golgi via dedicated transporters (Figure 6) (12). After entering the ER or the Golgi compartment, the donor substrates are transferred by the GT onto

their respective acceptor substrate, resulting in the release of the nucleotide. Nucleotide-monophosphate leaves the lumen in exchange with the incoming nucleotide-sugar (13). GDP-Man and UDP-Glc can also be converted to dolichol-phosphate-Man and dolichol-phosphate-Glc at the ER membrane.

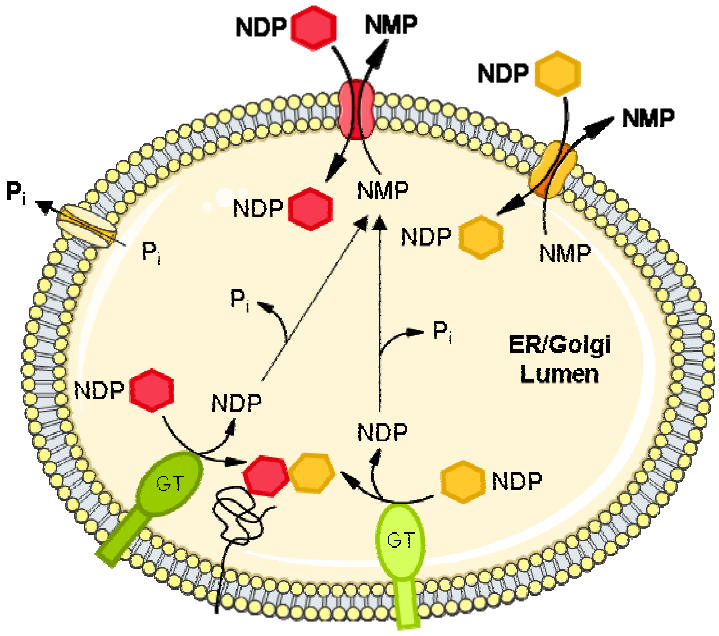


Figure 6: Nucleotide Sugar Transport Model. NDP-sugars are imported in the lumen of the ER/Golgi in exchange with NMP. In the lumen, the glycosyltransferase hydrolyzes the NDP-sugar. NDP is converted to NMP and inorganic phosphate (P_i) which leave the compartment.

The nucleotide transporters capable of transferring GDP-Fuc and UDP-Gal into the ER have not been characterized yet. This is controversial to the fact, that O-fucosylation of EGF domains in proteins (9) and galactosylation of hydroxylysines in collagen (see chapter 1.2.4.2) take place within the ER. Although no such transporter has been reported in the ER up to now (Table 1), there must be a mechanism to import GDP-Fuc and UDP-Gal into the ER as it was shown that these glycosylation steps take place in this compartment.

Table 1: Distribution of Currently Characterized Nucleotide Sugar Transporter in the ER and in the Golgi (according to (7, 12, 14))

Nucleotide Sugar Donor	ER	Golgi
CMP-Sia	-	+
GDP-Fuc	-	+
GDP-Man	-	+
UDP-Gal	-	+
UDP-Glc	+	+
UDP-GalNAc	+	+
UDP-GlcNAc	+	+
UDP-Xyl	+	+
UDP-GlcA	+	+

GTs are grouped into families according to their substrate- and linkage-specificities. The linkage catalyzed is usually mentioned in the name of the GT enzymes. For example a β(1-4)-galactosyltransferase catalyzes the formation of the glycosidic linkage in a β conformation of the C1 anomeric carbon of Gal with the C4 hydroxy group of the acceptor substrate.

1.1.4 Defects in the Human Glycome

The glycome, defined as all glycans synthesized in an organism, is as large as the proteome. Several glycosylation pathways are required for a normal development and physiology of an organism. Defects in glycosylation pathways have various impacts on biological functions. Accordingly, glycosylation defects are at the origin of several diseases in humans (Table 2).

A large family of diseases called congenital disorders of glycosylation (CDG) is linked to the N-glycosylation pathway. Two forms can be distinguished, CDG type-I with defects in the assembly of the LLO or in the transfer to the proteins while CDG type-II is caused by defects of N-glycan trimming within the ER or of glycan elongation within the Golgi apparatus (15, 16). More patients are affected by O-glycosylation linked diseases than N-glycosylation linked diseases. Diseases in the O-glycosylation pathway occur mainly in the O-xylose and the O-mannose glycosylation pathways. An example for an O-linked glycosylation disease is the Walker-Warburg Syndrome. This syndrome is characterized by mutations in the O-mannosyltransferase, which

transfers O-mannose on α-dystroglycan. Also the O-glycans on collagen may be important and it is likely that defects in the collagen glycosyltransferases could lead to an extracellular matrix disease (see also chapter 1.2.4.2).

Table 2: Representative Selection of Human Diseases Caused by Defects in Glycosylation Synthesis Pathways

Human Disease	Affected Enzyme/Pathway
CDG type-I	defect in LLO assembly or glycan transfer to proteins
CDG type-II	defects of N-glycan trimming in ER or elongation of glycan in Golgi
Mucolipidosis II and III	lack of mannose-6-phosphate
Congenital muscular dystrophies	O-mannose glycans on α-dystroglycan
Tn Syndrome	erythrocytes express Tn antigen
Paroxysmal nocturnal haemoglobinuria	defect in GPI anchor

1.2 Collagen

The glycoprotein collagen is the most abundant protein in the animal kingdom. It represents about one third of all proteins in the human body and is found in all tissues. Collagens are particularly abundant in bone, cartilage, tendon, ligaments, skin and the walls of blood vessels. By now, at least 44 genes coding for 29 types of collagen have been identified (17-19). Beside collagens, several other proteins, which do not belong to the collagen family, have a collagenous domain in their sequence.

Functionally, collagens are involved in the maintenance of the structure of different tissues, they are involved in cell adhesion and organogenesis or they regulate tissue repair in wound healing (20, 21). Even fragments of collagen, such as endostatin, can work as bioactive peptides (22).

1.2.1 Collagen Structure

The collagen molecule is composed out of three left handed α-chains, which coil around each other to form a right handed triple helix (Figure 9). Collagen molecules can consist of either three identical or different α-chains forming homotrimeric (e.g. collagens type II and III) or heterotrimeric structures (e.g. collagen type I), respectively. Collagens are characterized by at least one collagenous domain represented by the repetition of the triplet glycine-X-Y, where X and Y are often the amino acids proline and lysine. The presence of glycine in every third position is essential for the helical configuration, as glycine is the only amino acid which is small enough to fit into the centre of the helix. At their extremities collagen molecules contain non-triple helical domains, the N- and C- propeptides.

1.2.2 Collagen Nomenclature

Collagen types are numbered with Roman numerals in the order of their discovery. The α-chains are numbered with Arabic numerals with the collagen type indicated in parenthesis. For example, α1(III) stands for the α1 chain of type III collagen. The genes encoding collagen proteins are named by the prefix COL followed by the Arabic number for the chain, the ′A′ which stands

for the α-chain and the Arabic number for the chain. Accordingly, *COL1A2* encodes the α2 chain of collagen type I.

The collagen superfamily can be subdivided into two major classes based on their structure and supramolecular organization: the fibril-forming collagens and the non-fibrillar collagens (Figure 7).

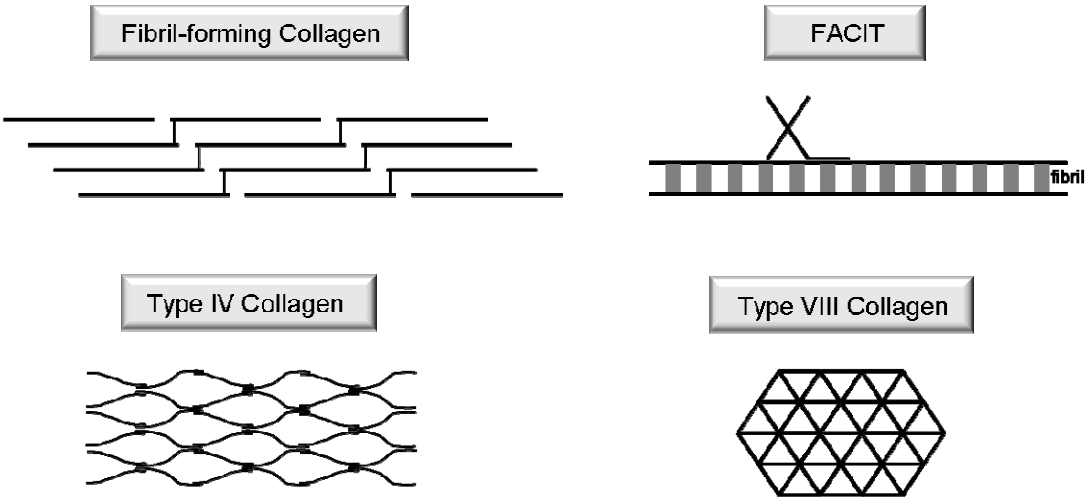


Figure 7: Molecular Organization of Collagens. Four representatives of the collagen superfamily are shown: fibril-forming collagen, FACIT, and the network-forming collagens, type IV and type VIII.

1.2.2.1 Fibril-forming Collagens

Fibril-forming collagens are the most abundant type of collagen, representing about 90 % of all collagens. The classical fibril-forming collagens comprise type I, II and III as well as type V and XI collagens. The latter are classified as fibril-forming collagens on the basis of their homology to collagen type I-III. The fibril-forming collagens are characterized by the ability to form supramolecular aggregates with a characteristic structure. They form 300 nm long unbranched fibrils with a triple helical domain of about 1000 amino acids, a highly conserved non-collagenous C-terminus and a variable non-collagenous N-terminus. At the electron microscopic level, a characteristic band pattern with a periodicity of about 67 nm, the so called D-period, is seen in the collagen fibrils (Figure 11) (20).

The molecules are synthesized as procollagens which are large precursor molecules containing large non-collagenous N- and C-propeptides linked to

the triple helical domain by the short non-collagenous telopeptides. These telopeptides are the sites where the intermolecular cross-linking for the collagen fibril stabilization takes place. This results in the formation of quarter-staggered structures (23). This collagen structure provides the mechanical strength of the body in the skeleton, skin, blood vessels, nerves and the intestines.

1.2.2.2 Non-fibrillar Collagens

Collagens that do not form collagen fibrils belong to the family of non-fibrillar collagens, which include type IV and type VI-X collagens for example. This group is largely heterogeneous in its structural diversity, tissue distribution, macromolecular organization and function. A common feature of these collagens is that non-collagenous sequences are interrupting the collagenous domain.

Non-fibrillar collagens are further divided in different subgroups. For example the FACIT, the fibril associated collagens with interrupted triple helices (e.g. types IX, XII, XIV, XVI, XIX), are characterized by short triple helical domains interrupted by non-collagenous sequences and associate as single molecules with large collagen fibrils. They can act as molecular couplers that organize and anchor fibrils with other extracellular matrix molecules (21). Furthermore, network-forming collagens, like the collagens type IV, VIII and X, form the essential scaffold of basement membranes. The transmembrane collagens (types XIII, XVII, XXIII, XXV) function both as cell surface receptors and as extracellular matrix components (24). Finally, multiplexins, such as collagen types XV and XVIII, contain multiple triple helical domains with several interruptions and are also found in basement membranes (21).

1.2.2.3 Non-collagenous Proteins

As mentioned previously, the collagenous domain is not confined to collagens, but is also found in several other proteins, such as the hormone adiponectin (ADIPOQ) (25), the mannose-binding lectin (MBL) (26), the C1q complement protein (27), the COLQ subunit of the acetylcholine esterase complex (28) and the surfactant proteins SP-A and SP-D (26).

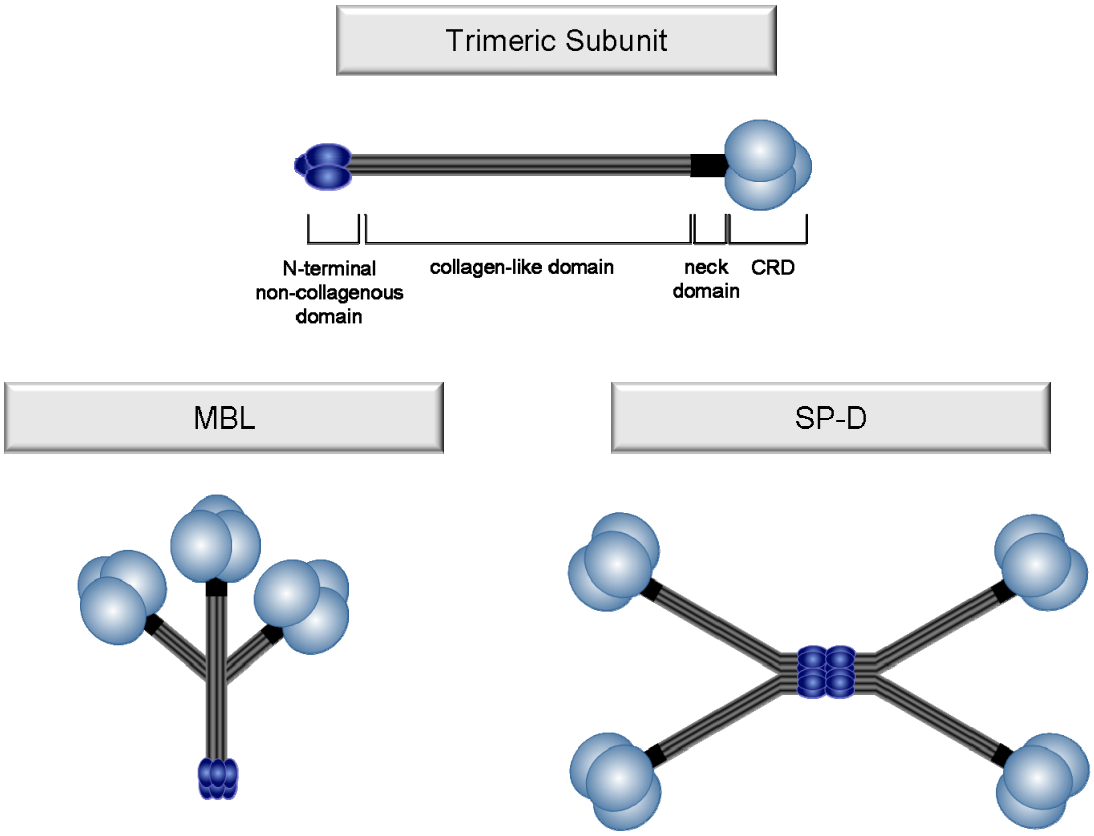


Figure 8: Non-collagenous Proteins. A schematic trimeric subunit with the N-terminal, collagen-like, neck and CRD domain is shown. Three MBL subunits associate and form the high oligomeric MBL molecule. Four subunits associate to form the SP-D molecule.

ADIPOQ is a hormone secreted from adipose tissue and is involved in Glc metabolism and fatty acid catabolism (29). It is composed of a signal peptide and three domains, namely a variable region, a collagenous domain and a globular domain. Through its collagenous domain, ADIPOQ can assemble in several oligomeric isoforms, the trimeric, hexameric and the high molecular weight (HMW) oligomeric complexes. The collagenous domain of ADIPOQ contains four hydroxylated lysine residues, which are glycosylated. These glycans play a crucial role in building the HMW complexes. Additionally, the hydroxylysines and the glycans might enhance the secretion of the HMW ADIPOQ (29).

MBL, SP-A and SP-D belong to the collagen containing C-type lectins called collectins. Their primary structure is organized in four domains: a short N-terminal, a long collagen-like, a neck and a carbohydrate recognition domain (CRD) (Figure 8). The cysteine containing N-terminus contributes to disul-

phide bond dependent oligomerization between monomers. The collagen-like domain forms a triple helix, which is associated with oligomerization and maintenance of the molecular structure. The neck domain forms a short trimeric coiled coil, which links the collagen-like domain to the CRD. Finally, the globular structure at the C-terminus comprising the CRD mediates calcium dependent ligand-binding (30, 31). All collectins have a mannose-type CRD (30, 32).

MBL is synthesized in the liver and is secreted into the blood stream where it has an important function in the innate immune system (33, 34). It binds to carbohydrates through its CRD. MBL is an oligomeric molecule made up of three polypeptides. Through the formation of a collagen triple helical structure, three polypeptides form a so called subunit. Several subunits can form higher oligomeric structures through disulphide bonds (35). In the case of MBL, three subunits are required to form the nascent protein (Figure 8). The lung surfactant proteins SP-A and SP-D are hydrophilic collagenous proteins (36). Lung collectins recognize and interact by their CRD with a wide range of carbohydrates on microbial surfaces (37). The collagen-like domain interacts with receptor molecules present on various immune cells to initiate clearance. In the case of SP-D, four subunits (Figure 8) form the nascent protein, while SP-A has an hexameric structure, in which six subunits oligomerize to form the nascent protein (36-38).

1.2.3 Collagen Biosynthesis

Collagen biosynthesis is a very complex process with the involvement of many intra- and extracellular proteins. Many of these associated proteins are unique to collagen.

Procollagen α-chains are synthesized on the ribosomes of the rough ER (Figure 9). As all of the collagen molecules are secreted, collagens are synthesized with a signal peptide which is cleaved within the ER lumen. After removal of the signal peptide, several proline and lysine residues are hydroxylated by three prolyl-4-hydroxylases (39), one prolyl-3-hydroxylase (40) and three lysyl hydroxylases (41, 42). 4-Hydroxyproline is essential for the folding and stability of the collagen helix. Hydroxylysines are involved in the formation of cross-links and they can also serve as acceptor sites for carbo-

hydrates which are added by a collagen galactosyl- und glucosyltransferase. Most of these modifications occur co-translationally, but continue post-translationally until triple helix formation prevents further modifications (43).

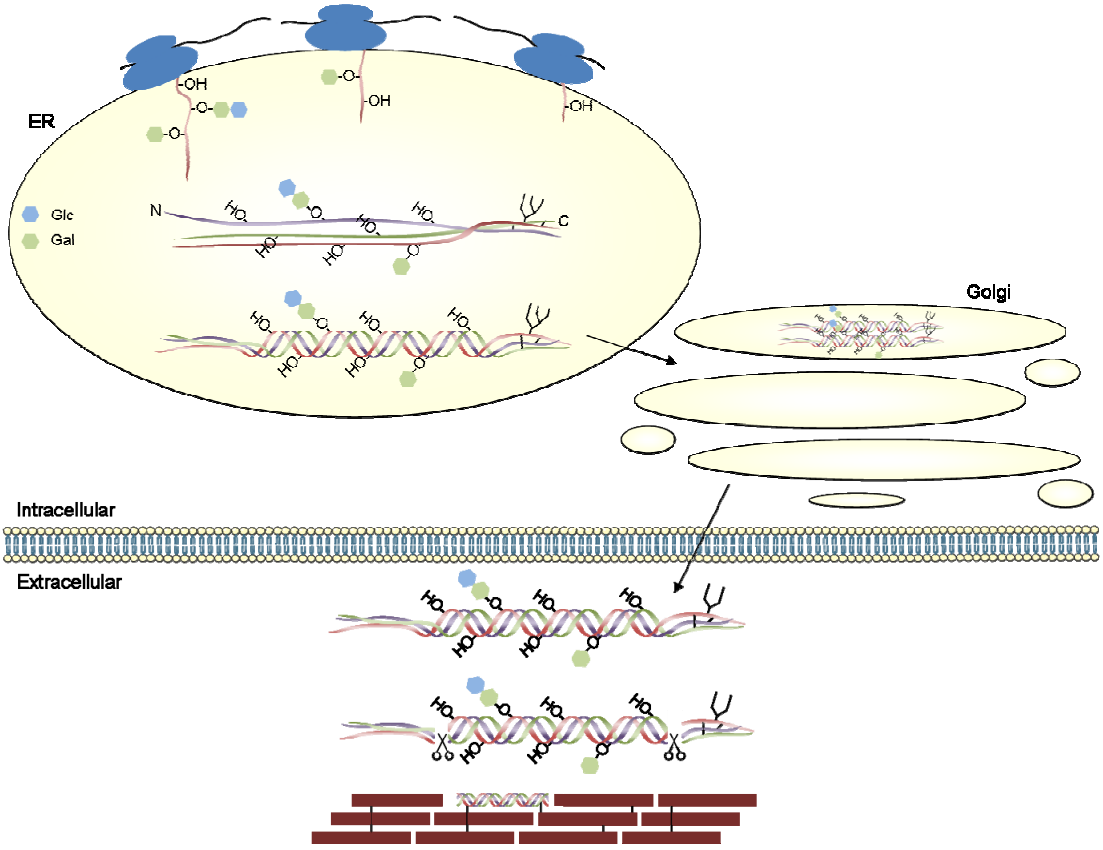


Figure 9: Collagen Biosynthesis: The polypeptides of the collagen molecule are synthesized on the ribosomes of the rough ER. In the ER lumen, the molecule is modified and forms a triple helix which is transported to the Golgi and which is finally secreted into the extracellular matrix where the propeptides are cleaved. Several of the resulting collagen molecules associate and form a fibril.

The folding of the procollagen into a triple helix is initiated by the assembly of three left handed α-chains at their C-propeptides after complete translocation into the lumen of the ER. Since the collagen types are defined on the basis of their composition with different α-chains, it is very important, that the correct three α-chains associate to form the fibril. This type specific assembly of individual α-chains is assured by the C-propeptides of the procollagen molecule. The association of the C-propeptides is stabilized by intermolecular disulphide bonds.

The 250 amino acids long C-propeptides are non-collagenous sequences, which show a high homology to each other. They all contain one N-glycan and eight cysteine residues. Four of these cysteine residues are located within the N-terminus of the C-propeptide and form intermolecular disulphide bonds, while the four C-terminal located cysteines form intramolecular disulphide bonds (44, 45) (Figure 10). The three α-chains begin to twist around each other propagating the formation of the triple helix towards the N-terminus (46). The propagation of the helix formation is done in a zipper-like manner. One single cell can simultaneously express several types of collagen. For example, skin fibroblasts synthesize the very homologous but genetically distinct collagens type I, III and V. It has to be assured that the different procollagen chains are able to discriminate from each other and assemble in a type specific manner. A discontinuous sequence of 15 hydrophilic amino acids in the C-propeptide has been shown to direct procollagen self-association in a type specific manner (47).

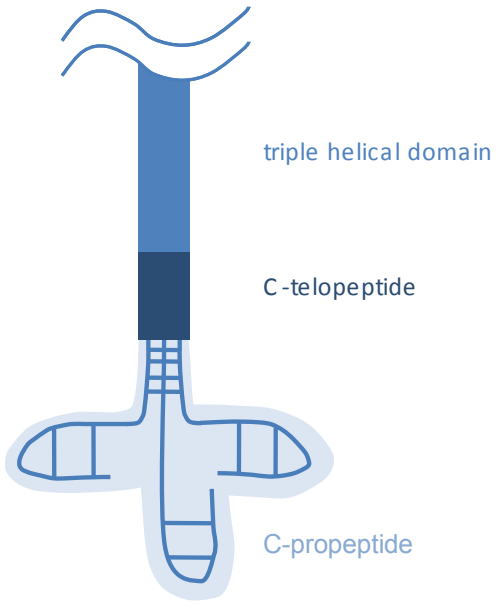


Figure 10: Schematic Representation of the C-propeptide Domain. Each polypeptide forms two intramolecular disulphide bonds in its C-terminal C-propeptide, whereas four intermolecular disulphide bonds are formed in the N-terminal region of the C-propeptide.

After completion of the triple helix formation, procollagen molecules are transferred to the Golgi apparatus, where they form aggregates by staggering and are secreted into the extracellular matrix (48-50). This transport of procollagen from the ER to the Golgi involves the chaperone HSP47, which

associates with the triple helical collagen in the ER and dissociates within the Golgi due to the pH shift. It has been suggested that HSP47 could stabilizes correctly folded collagen triple helix intermediates and to control lateral aggregation of collagen molecules (51, 52).

Currently, it is not fully resolved how procollagen molecules are transported through the secretory pathway. The size of the classical ER-Golgi transport vesicles is 60-80 nm. These vesicles seem to be too small to contain procollagen which has a length of 300 nm. Latest studies showed that the transport of procollagen shares some common features with well studied proteins as the vesicular stomatitis virus membrane protein ts-O45-G in ER-Golgi transport, but also differs considerably at the molecular level (53). It has been proposed that procollagen leaves the ER at sites devoid of COPII coat, although it appears that COPII is required for ER exit (53). Later, during transit through the Golgi, procollagen does not need to leave the lumen of the Golgi cisternae via vesicular carriers (54).

In the extracellular matrix, procollagen molecules are processed to collagen. The N- and C-propeptides are cleaved by propeptidases leading to the formation of the mature collagen molecule. Collagen self associates into higher fibrillar supermolecules in which each molecule is displaced about one-quarter of its length along the axis of the fibril (20, 55) (Figure 11). Finally, the collagen molecules are cross-linked between each other or within the collagen molecule.

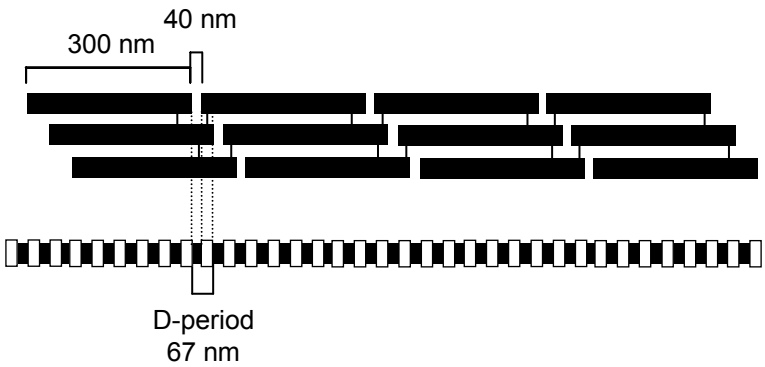


Figure 11: Supramolecular Assembly of Fibril-forming Collagen. In the quarter-staggered form each collagen molecule is displaced about one-quarter of its length to form a collagen fibril. Each fibril-forming collagen molecule has a length of 300 nm. The gap which separates consecutive molecules is 40 nm causing the characteristic appearance of the collagen fibril with a D-period of 67 nm. Intermolecular cross-links are shown by lines.

1.2.4 Collagen Modifications

The biosynthesis of collagen involves many posttranslational modifications that precede triple helix formation. Several ER resident specific enzymes are involved in this process.

1.2.4.1 Hydroxylation

Some proline and lysine residues of the collagen polypeptide are hydroxylated. This modification is only performed on collagen α-chains as further modifications are prevented by the triple helical structure of the final collagen (56-59). In the hydroxylation reaction, three different enzymes are involved: the lysyl hydroxylase (LH), the prolyl-4-hydroxylase (P4H) and the prolyl-3-hydroxylase (P3H). These three enzymes belong to the family of 2-oxoglutarate dioxygenases, which require Fe^{2+}, O_2, 2-oxoglutarate and ascorbate for their activity. In the reaction, 2-oxoglutarate is decarboxylated to succinate. One oxygen atom of O_2 is then incorporated into succinate, the second one being incorporated into the newly formed hydroxy group of hydroxylysine or hydroxyproline (Figure 12).

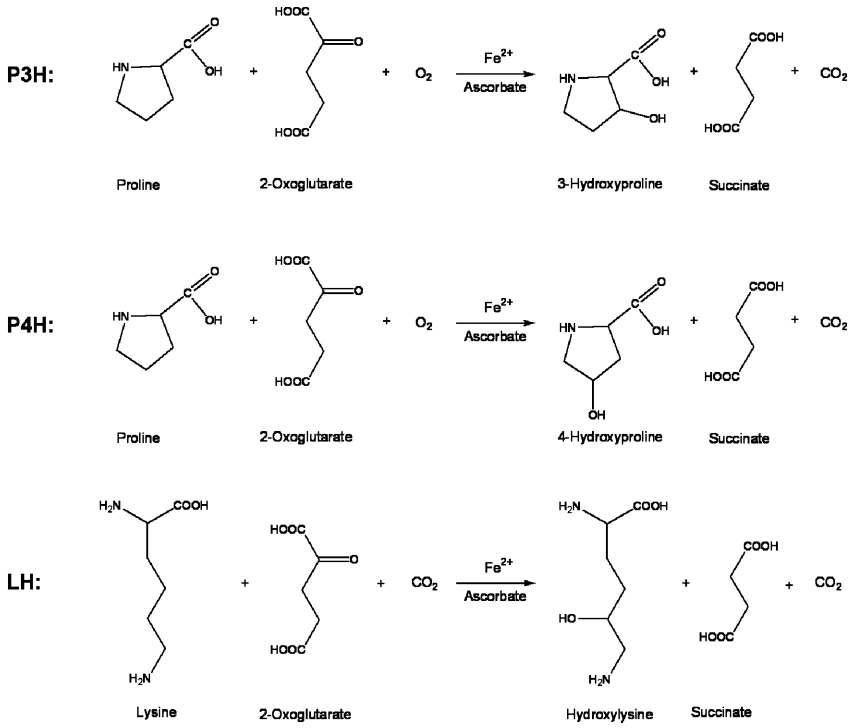


Figure 12: Collagen Hydroxylation: The three reactions of collagen hydroxylation are shown. These reactions are catalyzed by different collagen hydroxylases, the prolyl-3-hydroxylase (P3H), the prolyl-4-hydroxylase (P4H) and the lysyl hydroxylase (LH).

The catalytically critical Fe^{2+} binding residues are two histidines and one aspartate, which are conserved in all 2-oxoglutarate dioxygenases (60).

1.2.4.1.1 4-Hydroxyproline

The P4H catalyzes the formation of 4-hydroxyproline (4Hyp) in the sequence Gly-X-Pro in collagen and several other proteins with collagenous domains. *In vitro* the enzyme does not act on the single amino acid acceptor proline. It needs a minimal sequence requirement, which is fulfilled by the triplet -X-Pro-Gly-. It is absolutely necessary that proline is flanked by at least one amino acid on each side (61). 4Hyp stabilizes the collagen triple helix by H-bonds with water molecules, which is an essential process for the stability and folding of the collagen triple helix (57). These water-mediated H-bonds link the hydroxylgroup of 4Hyp with the carbonyl group of either glycine within the same chain or with 4Hyp in the adjacent chain (62).

The collagen P4H is a $\alpha_2\,\beta_2$ tetramer. The catalytical active site is located within the α subunit, while the β subunit contains a disulfide isomerase (PDI) activity. It was shown that the monomeric subunits do not possess any hydroxylase activity by themselves (59, 60, 63).

Three different isoenzymes exist for the α subunit, namely type-I ($[\alpha(I)]_2\,\beta_2$), type-II ($[\alpha(II)]_2\,\beta_2$) and type-III ($[\alpha(III)]_2\,\beta_2$) P4H (64-67). Type-I and type-II P4H show very similar but not identical enzymatic properties (65). Type-I P4H is the main form in most cell types, type-II is the major form in chondrocytes, osteoblasts and endothelial cells, while type-III is expressed in many human tissues, but at a much lower extent than type-I and type-II, respectively (64, 68, 69).

The β subunit with the PDI activity catalyzes the formation of disulfide bond formation in the C- and N-propeptides of procollagen. The main function of the PDI subunit associated with P4H is to keep the insoluble α subunit in an active, non-aggregated conformation and to retain the enzyme within the lumen of the ER (70).

The production of a stable P4H enzyme *in vivo* requires the co-expression of collagen polypeptides. This was demonstrated by expressing both subunits of the P4H in the yeast *Pichia pastoris*, which has no endogenous P4H activ-

ity. Most part of the recombinantly expressed P4H subunits were unassembled, a finding that could be prevented by coexpression of collagen type-III (71). A possible explanation could be that the tetramer becomes rapidly dissociated in its subunits in the absence of collagen. This dissociation prevents the consumption of 2-oxoglutarate, O_2 and ascorbate. Collagen synthesis appears to have a mutual control mechanism: stable production of P4H depends on collagen expression while stable collagen triple helix formation depends on proper P4H function.

1.2.4.1.2 3-Hydroxyproline

The collagen P3H catalyzes the formation of 3-hydroxyproline (3Hyp) in collagen. As 3Hyp is dependent on the presence of 4Hyp, 3Hyp is found in almost all collagens in the sequence Gly-3Hyp-4Hyp-Gly (72, 73). This suggests that the main substrate sequence for the P3H is Gly-Pro-4Hyp-Gly. The largest amount of 3Hyp is found in collagen types IV and V. 3Hyp is much less frequent than 4Hyp in the total amino acid content of collagen and its expression varies among the different collagen types (40).

P3H was originally reported as leprecan, a chondroitin sulphate proteoglycan (74) and later as growth suppressor GROS1 (75). Sequence analysis revealed three proteins which are closely related in human, mouse and chicken, called P3H1, P3H2 and P3H3. P3H1 is found in an intracellular collagen modification complex together with the cartilage-associated protein (CRTAP) and cyclophilin B (40, 76). It is found in tissues, in which fibril-forming collagens are expressed. The other two P3H family members P3H2 and P3H3 might have different substrate specificities. Based on an assay using labelled procollagen as a substrate, it was shown that the chicken P3H1 exhibits P3H activity (40) *in vitro*.

The function of 3Hyp is not well understood. Recent investigations have shown that P3H1 is crucial for bone development and collagen helix formation. Cabral *et al* (77) studied lethal to severe osteogenesis imperfecta-like bone dysplasia patients which do not show any collagen type I mutation. Their *P3H1* mRNA (also called *LEPRE1* mRNA) level was markedly decreased and the protein was absent. Different mutations in the *LEPRE1* gene were identified. The absence of P3H1 in these patients affects folding of col-

lagen type I. As folding is delayed, the molecule is exposed for a longer time to the lysyl hydroxylase and the P4H which results in over-modified collagen type I. The overmodification of the collagen may suggest that P3H1 functions as a collagen chaperone. These data delineate a new recessive bone disorder caused by deficiency of P3H1 called osteogenesis imperfecta type VIII (77, 78).

1.2.4.1.3 5-Hydroxylysine

LH catalyzes the formation of 5-hydroxylysine in collagen domains by the hydroxylation of lysines found in the Y position of the triplet G-X-Y. The non-helical C- and N-telopeptides with the sequence X-K-A or X-K-S are also hydroxylated on lysine residues (79). Hydroxylation of lysine residues is a typical modification of collagen. The extent of lysyl hydroxylation varies in different collagen types. Also, the same collagen type can be differently lysyl hydroxylated in different tissues. Hydroxylysines have two important roles: first they are essential for the formation of stable intra- and intermolecular cross-links and secondly they serve as acceptor sites for carbohydrates. Lysyl hydroxylation is only possible before the collagen triple helix is formed (72).

The LH family comprises the three different isoenzymes LH1, LH2 and LH3 (42, 80-82). Further variation occurs by alternative splicing of LH2, leading to a short and a long form of the enzyme, the LH2a and LH2b, respectively. The identity between LH1 and LH2 is 75 % while LH3 shares 59 % identity with both LH1 and LH2. Phylogenetic analysis revealed that the LH isoforms derive from one ancestral gene by two duplication events. LH1 and LH2 which share a higher level of sequence identity result from a more recent duplication than LH3 which appears to be the ancestral gene (83).

The LH1 isoform is expressed in many tissues such as skin, lung, placenta, cartilage, spleen, brain and liver (84). It was shown that the enzyme carries two N-glycans of which one is important as its mutation leads to a drastic reduction of the LH activity (85). Mutations in LH1 cause Ehlers Danlos Syndrome VIa (see chapter 1.2.5.1).

LH2 is a specific collagen telopeptide lysyl hydroxylase, whereas all three isoenzymes are able to hydroxylate lysine residues in the collagenous domain (86). As mentioned above, LH2a is the short and LH2b the long form.

LH2b has an additional exon which results in additional 21 amino acids. LH2a is the major form and expressed in several organs as pancreas, placenta, liver, heart, kidney and spleen (80, 87). Mutations in the gene coding for LH2 have been associated with the Bruck syndrome (see chapter 1.2.5.3).

The LH3 enzyme is mainly expressed in the heart, pancreas and the placenta (42). Up to now no disease could be attributed to mutations in the *LH3* gene located on chromosome 7q36 (42, 88). LH3 knock-out mice die at embryonic day 9.5 (89). Unlike LH1 and LH2, LH3 has been reported to be a triple active enzyme and to exhibit beside its LH activity also two additional enzyme functions, namely a collagen galactosyl- and glucosyltransferase activity (90, 91). These activities have been mapped to the N-terminal domain of the enzyme while the LH activity is located in the C-terminus.

Like the prolyl hydroxylases, the lysyl hydroxylases belong to the 2-oxoglutarate dioxygenase family and require Fe^{2+}, O_2, 2-oxoglutarate and ascorbate. The active LH enzyme is a homodimer and is a peripheral membrane protein located in the ER lumen, although no ER retrieval signal is found (82, 92). Unlike the soluble prolyl hydroxylases, LH is membrane bound (93).

1.2.4.2 Collagen Glycosylation

The O-glycosylation of collagen has been first described by Grassmann and Schleich in 1935 (94). Collagen carries either the monosaccharide Gal or the disaccharide Glc-Gal which are glycosidically linked to the C5-hydroxygroup of hydroxylysine. The structure of the glycan was determined by Spiro in 1967 as being Glc(α1-2)Gal(β1-O)Hyl (95) (Figure 13). Gal and Glc are transferred by either the collagen galactosyltransferase (ColGalT) or the collagen glucosyltransferase (ColGlcT), respectively. The ColGalT transfers Gal from UDP-Gal to the hydroxylysine of procollagen but not to free hydroxylysine (96). ColGlcT transfers Glc from UDP-Glc to galactosylated hydroxylysine. These reactions are dependent on the bivalent cofactor Mn^{2+} (97).

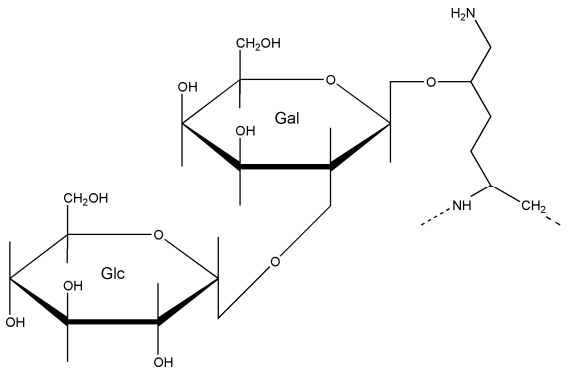


Figure 13: Glycan of Collagen: Structure of the Glc(α1-2)Gal(β1-O) disaccharide attached to a hydroxylysine residue of collagen.

Since glycosylated hydroxylysines were found on the nascent collagen polypeptide chain, it was concluded that collagen glycosylation already starts co-translationally (98, 99). Collagen glycosylation ceases after completion of the triple helix formation. This indicates that only collagen α-chains, but not triple helical collagen can serve as a substrate for the ColGalTs (100, 101). A free ε-amino group of the hydroxylysyl residue and a nonhelical polypeptide conformation are absolute requirements for both collagen glycosyltransferases (21).

The extent of collagen glycosylation varies in different types of collagen, but also in the same type of collagen in different tissues (90). It is known that type IV collagen is the most glycosylated collagen, whereas collagen type I carries only very little amount of carbohydrates (97). The function of these carbohydrates is unknown. Carbohydrates appear to influence both the lateral packing of the collagen molecule into fibrils and the diameter of the fibrils in fibril-forming collagens (102). The collagen domain of ADPIOQ and MBL also carry glycosylated hydroxylysine residues, which are important for the oligomerization and proper secretion of these proteins (25, 103).

Myllylä and coworkers described in the years 2000 and 2002 for the first time that the collagen lysyl hydroxylase LH3 shows low glucosyl- (91) and galactosyltransferase (90) activities. However these ColGalT and ColGlcT activities are so low that they might have little biological significance (104). There is further evidence that there must exist two additional enzymes responsible for most of the collagen glycosylation *in vivo* as previous data showed that the two reactions are catalyzed by two different enzymes in vertebrates (63, 100, 105). In LH3, the catalytically active site of the collagen glycosyltransferase is

comprised within the N-terminal region of the protein, whereas the LH activity is mediated by the C-terminal region of the protein (104). The glycosyltransferase domain contains a DxD motif which is known to bind to divalent metal ions and to the hydroxyl group of the ribose moiety. Mutating the motif causes inactivation of the collagen glycosyltransferase activities of the LH3 enzyme (90, 106).

LH3 knock-out and LH3 transgenic mice with a mutated LH activity domain but with intact collagen glycosyltransferase activities were generated (107). These mice serve as a good model to study the importance of collagen glycosylation linked with the LH3 triple active enzyme. By studying the LH3 knock-out embryos it was observed that the total collagen amount (both intra- and extracellular) decreased. Due to the impaired secretion machinery, collagen type IV aggregated which lead to a dilatation of the ER and to a condensed Golgi. The secretion of the normally highly glycosylated collagen type IV was further studied in fibroblasts of the knock-out mice. Collagen type IV aggregates within the ER due to deficient glycosylation and is rapidly degraded within the cell which leads to a lower amount of collagen type IV compared with wildtype fibroblasts. Sipilä *et al* suggested that collagen type IV might aggregate due to incorrect folding, which could result due to the lack of glycans (108). The highly glycosylated network-forming collagen type VI is unique among collagens as it forms tetramers already inside the cells (109). In knock-out fibroblasts no tetramers could be detected demonstrating the important function of the glycans in collagen type VI tetramerization. However, in fibroblasts of LH3 transgenic mice with a mutated LH activity domain, normal tetramerization and secretion but abnormal aggregation and distribution of type VI collagen was detected (108). These results suggest that collagen glycosylation plays an important role in collagen folding and quality control of the proteins as seen for the N-glycans. Furthermore, the glycans appear to be involved in the formation of supramolecular collagen structures.

1.2.4.3 Collagen Cross-linking

The mechanical strength of collagen fibrils depends on the highly regulated mechanism of covalent intermolecular cross-link formation. Collagens are crosslinked through a reaction of aldehydes, which are generated from lysine

or hydroxylysine residues by the extracellular enzyme lysyl oxidase and the lysyl oxidase-like enzymes (110). The cross-links connect non-helical telopeptides with helical regions in adjacent collagen molecules. The formation of the cross-link begins with a step catalyzed by the lysyl oxidase, which oxidatively deaminates the ε-amino group of the telopeptide lysine or hydroxylysine to the corresponding reactive lysylaldehyde and hydroxylysylaldehyde, respectively. The subsequent steps occur spontaneously. The aldehydes are involved in the formation of various cross-links, either by aldol condensation of two aldehydes or by condensation of an aldehyde and an ε-amino group of a second lysine, hydroxylysine or glycosylated hydroxylysine. Hydroxylysine-derived cross-links are more stable than those formed by lysine aldehydes. The lysine aldehyde pathway occurs mainly in adult skin, cornea and sclera, whereas the hydroxylysine aldehyde pathway occurs in bone, cartilage, ligament and embryonic skin (111).

1.2.5 Diseases Related to Collagen Biosynthesis Defects

Collagens are susceptible to acquire mutations which can lead to a variety of diseases of bone, cartilage or blood vessels. This is due to the high amount of invariant and essential glycine residues. Substitution of either of the first and second nucleotide positions in the codon for glycine GGN leads to an amino acid exchange on the protein level. This exchange leads to a change of the triplet G-X-Y, which is fatal, as no other amino acid will fit into the centre of the collagen helix. Additionally, collagens contain many exons which could be skipped while the gene is processed (21).

According to the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) (http://www.niams.nih.gov/) there are currently more than 200 heritable disorders of the connective tissue linked with mutations in collagen or collagen processing genes (112).

Table 3: Examples for Human Collagen Diseases

Affected Gene	Human Disease
COL3A3, COL4A4, COL4A5, COL4A6	Alport Syndrome
COL6A1, COL6A2, COL6A3	Bethlem Myopathy
LH2	Bruck Syndrome
COL5A1, COL5A2	Ehlers Danlos Syndrome Type I,II
COL3A1	Ehlers-Danlos Syndrome Type IV
COL18A1	Knobloch Syndrome
COL1A1, COL1A2	Osteogenesis Imperfecta
P3H1	Osteogenesis Imperfecta Type VIII

1.2.5.1 Ehlers Danlos Syndrome

The family of Ehlers Danlos Syndromes (EDS) is a heterogeneous group of connective tissue disorders which are clinically characterized by fragility of the skin, hypermobility of the joints and hyperextensibility. The syndrome is named after the Danish dermatologist Edvard Ehlers and the French dermatologist Henri-Alexandre Danlos (113). The disorder can be inherited autosomal dominantly, autosomal recessively or X-linked according to their different subtypes. Based on the “Villefranche Nosology, 1997” six major classes of EDS are known based primarily on the molecular defects (114). The estimated prevalence is 1 in 5′000 individuals depending on the type of EDS (115). 90 % of all EDS cases represent EDS type I or type II and are therefore called classical type. The defect is caused by mutations in the collagen type V gene. The cause of type III is unknown. EDS type IV is caused by mutations in the collagen type III gene while mutations in the collagen type I gene lead to EDS type VIIa and VIIb. These types of EDS are autosomal dominant. Another subtype of EDS type VII, named EDS type VIIc is caused by the mutation of a collagen processing enzyme, the N-propeptidase. EDS type VI, also called kyphoscoliotic type of EDS, is an autosomal recessive disease which is biochemically divided into two subgroups: type VIa: with a low activity of LH1, due to mutations in the *LH1* gene, and type VIb with a normal LH activity. Up to now, the cause of EDS VIb is not known. The major

symptoms of patients suffering from EDS type VI are progressive kyphoscoliosis, muscular hypotonia, joint laxity, scleral fragility and rupture of the ocular globe (21).

Table 4: EDS Types and their Causes

EDS Type	New Nomenclature	Affected Gene
EDS Type I	Classical Type	*COL5A1, COL5A2*
EDS Type II	Classical Type	*COL5A1, COL5A2*
EDS Type III	Hypermobile Type	COL3A*1, TNXB* (tenascin-XB)
EDS Type IV	Vascular Type	*COL3A1*
EDS Type V	Other Types	X-linked, cause unknown
EDS Type VI	Kyphoscoliotic Type	VIa: *LH1* VIb: cause not known
EDS Type VII	Arthrochalasic Type	VIIa, VIIb: *COL1A1, COL1A2*
EDS Type VIIc	Dermatosparactic Type	*ADAMTS2* (N-propeptidase of collagen type I and II)

1.2.5.2 Osteogenesis Imperfecta

Osteogenesis Imperfecta (OI), also known as brittle bone disease is a autosomal dominant genetic disorder. It is mainly characterized by bone fragility and fractures. The disease can be accompanied by reduced life span, bone deformities, blue sclera, hearing loss and other evidence of connective tissue abnormalities. The phenotypic presentation varies from mild to lethal. For the classification of OI, a scheme developed by Sillence and colleagues is used. The patients are classified by radiographic, genetic and clinical criteria and are assigned to one of the four groups, namely type I to IV (112). Three new uncommon types of OI have been described and assigned as OI type V to type VII (116-118). The cause of these OI types is unknown, but they do not show any mutations in the genes coding for COL1A1 and COL1A2. As type V to type VII show OI phenotype but no collagen type I mutation, it has been proposed to call them “syndromes resembling OI” (119).

Mutations in COL1A1 and COL1A2 account for 85 % of all OI cases. More than 250 distinct mutations are known. Most of them are single nucleotide

substitutions that exchange a glycine to another amino acid, thus resulting in a reduced collagen stability (120). The mildest form of OI (type I) is caused by mutations that generate a premature stop codon within the COL1A1 gene. Due to the silencing of one allele, the total amount of collagen is diminished as the transcripts of the altered collagen are unstable and undergo nonsense mediated RNA-decay (112).

1.2.5.3 Bruck Syndrome

The Bruck syndrome (BS) is an autosomal recessively inherited disorder. The disease is characterized by osteoporosis, joint contractures at birth, short stature and fragile bones. Often seen are also recurrent fractures of bones in infancy and early childhood, features reminiscent of OI (21). The biochemical analysis of bone in BS patients revealed an underhydroxylation of the telopeptides of collagen type I, but normal hydroxylation in the triple helix. LH2 is the lysyl hydroxylase which hydroxylates lysine residues in the telopeptide region of the collagen molecule. BS type II is defined by patients having mutations in the LH2 found on chromosome 3q23-q24 while BS type I is linked with mutations within chromosome 17p12 (86, 121, 122). This classification is based on genetic data as there are no phenotypic differences (86, 122).

1.2.5.4 Scurvy

Scurvy was a very common disease among sailors. The cause of scurvy is the inability of human beings to synthesize ascorbic acid within the liver because the last enzyme in the cascade to convert Glc to ascorbic acid is missing. Therefore, ascorbic acid has to be taken up by nutrition. Insufficient intake of ascorbic acid will result in scurvy. All mammals except Man, some monkeys and guinea pigs, are able to synthesize ascorbic acid in their liver (123). Only the above listed animals will develop scurvy if deprived of ascorbic acid in their diet.

Insufficient intake of ascorbic acid results in an impaired collagen biosynthesis. As ascorbic acid is a cofactor of the prolyl- and lysyl hydroxylases, the absence of ascorbic acid and thus the absence of hydroxyproline and hydroxylysine leads to unstable collagen polypeptides which are not able to self-assemble into triple helices. These instable collagen molecules lead to

haemorrhage of the gum and thus loss of teeth, poor wound healing, bruising and hyperkeratosis (124, 125). Scurvy can be prevented by a diet which includes citrus fruits and vegetables which are good sources of ascorbic acid.

2 Results

2.1 Characterization of the Human ColGalT

2.1.1 Manuscript 1

Core glycosylation of collagen is initiated by two β(1-O)galactosyltransferases

under revision in Molecular and Cellular Biology

Belinda Schegg, Christoph Rutschmann, Andreas J. Hülsmeier,

Thierry Hennet

Institute of Physiology and Center for Integrative Human Physiology,
University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich

Own Contribution

Affinity Chromatography

MS sample preparation

Cloning and Baculovirus protein expression

Enzymatic activity assays

HPLC analysis

Assay product identification

Northern Blotting

Deglycosylation of proteins

Core glycosylation of collagen is initiated by two β(1-O)galactosyltransferases

Running Title: Collagen galactosylation

Belinda Schegg, Christoph Rutschmann, Andreas J. Hülsmeier, Thierry Hennet§

Institute of Physiology and Center for Integrative Human Physiology, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich

§Corresponding author: Thierry Hennet
Institute of Physiology, University of Zurich,
Winterthurerstrasse 190, CH-8057 Zurich, Switzerland
Tel: +41-1-635-5080, Fax: +41-1-635-6814
E-mail: thennet@access.uzh.ch

Word counts: 1354 words in the Materials and Methods
2033 words in Introduction, Results and Discussion

Abbreviations: CEECAM1, cerebral endothelial cell adhesion molecule-1; ColGalT, collagen galactosyltransferase; ER, endoplasmic reticulum; Hyl, hydroxylysine; GHyl, galactosylated hydroxylysine; GGHyl, glucosyl-galactosylated hydroxylysine; MBL, mannose-binding lectin

Keywords: extracellular matrix, glycosylation, hydroxylation, mass-spectrometry, endoplasmic reticulum

Data deposition: The nucleotide sequences reported in this paper correspond to the GenBank/EBI Data Bank entries with the accession numbers NM_024656, NM_015101 and NM_016174.

Abstract

Collagen is a trimer of three left-handed alpha-chains representing repeats of the motif Gly-X-Y, whereas (hydroxy)proline and (hydroxy)lysine residues are often found at positions X and Y. Selected hydroxylysine are further modified by the addition of galactose and glucose-galactose units. Collagen glycosylation takes place in the endoplasmic reticulum before triple helix formation and is mediated by β(1-O)galactosyl- and α(1-2)glucosyltransferase enzymes. We have identified two collagen galactosyltransferases using affinity chromatography and tandem-MS protein sequencing. The two collagen β(1-O)galactosyltransferases corresponded to the GLT25D1 and GLT25D2 proteins. Recombinant GLT25D1 and GLT25D2 enzymes showed a strong galactosyltransferase activity towards various types of collagen and towards the serum mannose-binding lectin MBL, which contains a collagen domain. Amino acid analysis of the products of GLT25D1 and GLT25D2 reactions confirmed the transfer of galactose to hydroxylysine residues. The GLT25D1 gene is constitutively expressed in human tissues, whereas the GLT25D2 gene is only expressed at low levels in the nervous system. The GLT25D1 and GLT25D2 enzymes are similar to CEECAM1, to which we could not attribute any collagen galactosyltransferase activity. The GLT25D1 and GLT25D2 genes now allow addressing the biological significance of collagen glycosylation and the importance of this post-translational modification in the etiology of connective tissue disorders.

Introduction

Collagens are the most abundant proteins in the human body. To date, 29 types of collagen have been described, which are encoded by at least 44 genes (20, 36, 44). Collagens are characterized by domains representing repeats of the triplet Gly-X-Y, where proline and lysine are often found at the positions X and Y. The Gly-X-Y repeats are not confined to collagens, but are also found in several proteins, such as the hormone adiponectin (28), the mannose-binding lectin (MBL) (10), the C1q complement protein (34), the COLQ subunit of the acetylcholine esterase complex (4) and the surfactant proteins SP-A and SP-D (10).

After synthesis in the endoplasmic reticulum (ER), three procollagen subunits associate to build a right-handed triple helix. However, before the formation of the triple helix structure, the nascent procollagen polypeptides undergo several post-translational modifications. These modifications comprise the hydroxylation of selected proline (19) and lysine (32) residues, which are catalyzed by three prolyl-4-hydroxylases (16), one prolyl-3-hydroxylase (45) and three lysylhydroxylases (42). Hydroxylysine can be further modified by the addition of the monosaccharide Gal(β1-O) or the disaccharide Glc(α1-2)Gal(β1-O) (38).

Whereas the glycosylation of collagen has been first described by Grassmann and Schleich in 1935 (9) and the structure of the glycan determined by Spiro in 1967 as being Glc(α1-2)Gal(β1-O)Hyl (39), the molecular nature of the collagen glycosyltransferase enzymes has remained elusive up to now. Col-

lagen galactosyltransferase (ColGalT) and glucosyltransferase activities have been characterized using partially purified proteins (23, 30, 31), which appeared to be instable. Recently, the lysylhydroxylase-3 LH3 enzyme has been shown to catalyze a modest galactosyl- and glucosyltransferase activity, suggesting that this enzyme is a combined hydroxylase and glycosyltransferase (11).

Prolyl and lysyl hydroxylation contribute to the stability of the collagen triple helix whereas hydroxylysine is essential for the cross-linking of collagen molecules, thus ensuring the strength of collagen fibrils (27). By contrast, the biological significance of collagen glycosylation is still unclear. The collagen domain of adiponectin and mannose-binding lectin also carry glycosylated hydroxylysine residues, which appear important for the oligomerization and proper secretion of these proteins (6, 28).

The importance of collagen post-translational modifications is reflected by the diseases caused by defective collagen modifying enzymes. Mutations of the *LH1* lysylhydroxylase-1 gene lead to the connective tissue disorder Ehlers-Danlos syndrome type VI (13) and mutations in the *LH2* lysylhydroxylase-2 gene lead to the Bruck syndrome (43). The deficiency of the prolyl-3-hydroxylase-1 gene causes a severe form of osteogenesis imperfecta (5). The availability of the collagen glycosyltransferase genes will enable the comprehensive investigation of this post-translational modification in cellular and animal models and possibly in human diseases.

Materials and Methods

Affinity chromatography. Collagen glycosyltransferases were enriched by affinity chromatography as described by Myllyla *et al.* (1, 21, 23). Briefly, 10-day-old chicken embryos were homogenized in 225 mM mannitol, 75 mM sucrose, 50 µM DTT and 50 mM Tris-HCl, pH 7.4 at 4 °C and centrifuged at 15,000 x *g* for 40 min. Supernatants were filtered and proteins precipitated in 60 % $(NH_4)_2SO_4$. The pellets obtained after 20 min centrifugation at 15,000 x *g* were dissolved in 0.2 M NaCl, 50 µM DTT, 1 % glycerol, 20 mM Tris-HCl, pH 7.4 and dialyzed overnight against 2.5 l of enzyme buffer (0.15 M NaCl, 10 mM $MnCl_2$, 50 M DTT, 1 % glycerol, 50 mM Tris-HCl, pH 7.4). The chicken protein extracts were loaded on a column of agarose-bound bovine achilles collagen type I fragments as described previously (31). The column was washed with 5 volumes of enzyme buffer containing 500 µM UDP followed by elution with 0.1 % acetic acid. Collected fractions were immediately neutralized with 1 M Tris pH 8.0.

MS peptide analysis. The eluted fractions from the affinity chromatography were desalted and concentrated with Amicon Ultra 10 cartridges (Millipore). Portions of 2 µg protein were reduced in 0.6 M Tris pH 8.5, 50 mM DTT for 5 min at 80 °C and alkylated for 40 min at RT in the dark by the addition of iodoacetamide (Sigma-Aldrich, final concentration 200 mM) and desalted by adding 9 volumes of ice cold methanol for 18 h on ice. Alkylated proteins were digested for 18 h at 37 °C with 0.01 µg trypsin (Roche). ZipTip (Milli-

pore) purified peptides were then analyzed by liquid chromatography-MS. The desalted peptide digest was adjusted to 0.2 % formic acid, 3 % ACN and directly injected onto a custom packed 80 mm x 0.075 mm ProntoSil-Pur C18-AQ, 3 µm, 200 Å, column (Bischoff GmbH, Leonberg, Germany), connected to a LTQ-ICR-FT mass spectrometer (Thermo Scientific, Bremen, Germany). The peptides were eluted with a binary gradient of solvents A (3 % ACN, 0.2 % formic acid) and B (80 % ACN, 0.2 % formic acid) using an Eksigent-Nano-HPLC system (Eksigent technologies, Dublin, USA). The column was flushed for 16 min at a flow rate of 500 nl/min with 100 % buffer A. Buffer B was increased to 3 % over 5 min, to 60 % over 50 min, to 100 % over 3 min and held at 100 % for 7 min. During gradient elution, the flow rate was maintained at 200 nl/min. The mass spectral data were acquired in the mass range of 300-2000 m/z. Data dependent MS/MS spectra were recorded of up to four of the most intense ions with higher charge state than 1+ using collision-induced dissociation (CID). Target ions already selected for MS/MS were dynamically excluded for 60 s. Peptide signals exceeding 500 counts were subjected to CID with normalized collision energy of 32 %. MS and MS/MS data were searched using Mascot Server 2.1 (Matrix Science, London, UK) as the search engine. Modifications used include carbamidomethylation (Cys, fixed) and oxidation (Met, variable). The monoisotopic masses of +2, and +3 charged peptides were searched with a peptide tolerance of 2 ppm and MS/MS tolerance of 0.8 Da. MS/MS spectra were searched against the UniRef100 20051018 database (2764545 sequences;

1015909965 residues) downloaded from the European Bioinformatics Institute (http://www.ebi.ac.uk/uniprot/database/download.html) and the *Gallus gallus* predicted proteins database from the Ensemble Genome Browser ftp://ftp.ensembl.org/pub/current_chicken/data/fasta/pep/Gallus_gallus.WASHUC1.jul.pep.abinitio.fa.gz,August 2005).

Cloning and protein expression. The *GLT25D2*, *LH3* and *MBL* cDNAs were purchased from the RZPD repository (Berlin, Germany). The *GLT25D1* and *cerebral endothelial cell adhesion molecule-1 (CEECAM1)* cDNAs were cloned by RT-PCR from human fibroblast total RNA using the primers 5'-ATCTGAATTCCCTTTAAGGCGCGGCCAGAGTC-3', 5'-ATGTCTAGATGGAGCCTGGGCCACCGATG-3' for *GLT25D1*, and 5'-CGTAGAATTCGAGAGCTCCGGGGGCCGCT3', 5'-GACTATCTAGAGTAGTGGCCTGCTCCTGGAC-3' (Microsynth, Switzerland) for *CEECAM1.* The RT-PCR products were subcloned as *EcoR*I-*Xba*I fragments into the pFastBacI baculovirus transfer vector (Invitrogen). The *MBL* cDNA was subcloned into the *EcoR*I site of the pFmel-protA vector (47) to yield a protein-A fusion protein. The corresponding 732 bp *MBL* fragment was amplified with the primers 5'-ATCGAATTCATGGTGGCAGCGTCTTACTC-3' and 5'-ATCGAATTCAGGAGGGCCTGAGTGATATG-3'. Recombinant baculoviruses were produced in *S. frugiperda* Sf9 cells as described previously (12). Protein-A tagged MBL was co-expressed together with LH3, purified from the supernatant of infected Sf9 cells by IgG Sepharose chromatography (47) and subsequently used as accep-

tor for the enzymatic activity assay. The expression of the recombinantly expressed enzymes was analyzed on a 10 % Coomassie SDS-PAGE gel. Prior to loading on the gel proteins were enriched by concanavalin-A lectin Sepharose beads (Amersham). Protein bands were excised from the Coomassie SDS-PAGE gel, in gel digested with trypsin according to Shevchenko *et al* (33) and analyzed by LC-MS (procedure see MS peptide analysis section).

Preparation of ColGalT acceptors. Achilles collagen type I, bovine nasal septum collagen type II, human placenta collagen type III, IV and V (Sigma) were deglycosylated by trifluoromethane sulfonic acid (TFMS)-mediated cleavage (7, 37). Acceptor proteins (50 µg) were lyophilized followed by an incubation in a dry ice/ethanol bath for 20 min. Proteins were dissolved in 50 µl TFMS/toluene (16.6:1, v:v) (Sigma-Aldrich). Reactions were subsequently incubated at -20 °C for 24 h, then neutralized with 150 µl pyridine/H_2O (2:1, v:v) in the dry ice/ethanol bath followed by 15 min incubation on ice. The sample was mixed with 400 µl 50 mM ammonium acetate and dialyzed overnight against 2.5 l of 50 mM ammonium acetate.

Collagen glycosyltransferase assays. Baculovirus infected Sf9 cells were lyzed in 1 % Triton-X100 / TBS pH 7.4 for 10 min on ice and postnuclear supernatant was used as enzyme source. Collagen was heat denatured for 10 min at 60 °C in sodiumacetate pH 6.8 and rapidly cooled down to 0 °C before use. Assays were performed with 10 µl of Sf9 postnuclear supernatant

in a final volume of 100 µl containing 0.5 mg/ml collagen acceptors, 60 µM UDP-Gal or UDP-Glc, 50,000 cpm UDP-[^{14}C]Gal or UDP- [^{14}C]Glc (GE Healthcare), 10 mM $MnCl_2$, 20 mM NaCl, 50 mM MOPS pH 7.4, 1 mM DTT. Reactions were incubated for 3 h at 37 °C and stopped by the addition of 500 µl of ice cold 5 % TCA / 5 % phosphotungstic acid. Enzymatic activity assays for Km analysis were performed as described above but with different amounts of either collagen or UDP-Gal as a substrate for the reaction.

Amino acid analysis. The reaction products of the collagen galactosyltransferase assays were hydrolyzed in 4 M NaOH for 72 h at 105 °C and the resulting single amino acids were derivatized with Fmoc according to Bank *et al.* (2). RP-HPLC (LaChrom Hitachi, Merck) of single amino acids was performed on a ODS Hypersil column, 150 x 3 mm, 3 µm particle size (Thermo Electron Corporation) at 40 °C. The galactosylated Hyl (GHyl) and galactosyl-glucosylated Hyl (GGHyl) standards were kindly provided by Ruggero Tenni (University of Pavia) (41). Amino acids were separated at a flow rate of 0.2 ml/min using a gradient elution (Table A2) with the solvents 0.5 M citric acid, 5 mM $(CH_3)_4NCl$, pH 2.85 (A); 80 % of 20 mM sodium acetate trihydrate, 5 mM $(CH_3)_4NCl$, pH 4.5, 20 % of methanol (B); 100 % of ACN (C). Radiolabelled [^{3}H]Val and [^{14}C]Tyr (Moravek Biochemicals and Radiochemicals, USA) were used as internal standard. Radioactivity was counted in a β-counter (Tri-Carb 2900TR, Packard).

Northern blotting. The Northern Blot probes were synthesized by PCR using the following primers: GLT25D1 upper: 5'-GATGAGGCCGAGAGCTTC-ATGC-3', lower: 5'-GCATGAAGCTCTCGGCCTCATC-3' (Microsynth, Switzerland), giving a product of 676 bp; GLT25D2 upper: 5'-AAGCAGGCATC-CAGATGTACC-3' lower: 5'-TCCAGCTGAGCCTGGTCAATG-3' giving a product of 559 bp; CEECAM1 upper: 5'-GTGGATGGCTGGATGCTCAAC-3' lower: 5'-GACTATCTAGAGTAGTGGCCTGCTCCTGGAC-3' giving a product of 785 bp. *GLT25D1, GLT25D2* and *CEECAM1* cDNA probes were labeled with α[^{32}P]dCTP (Hartman Analytic, Germany) by random priming (Stratagene). Multiple human tissue RNA arrays (MTE array 3, BD Bioscience and First Choice Northern Human Blot 1, Ambion) were prehybridized with QuikHyb® hybridization solution (Stratagene) containing 100 µg/ml ultra pure herring sperm DNA (Invitrogen) for 1 h at 65 °C, then hybridized with 5 x 10^5 cpm of each labeled probe overnight at 65 °C. The arrays were washed in 0.1 x SSC, 0.1 % SDS up to 60 °C and exposed on BioMax XAR film (Kodak) for 24 h at -80 °C.

Results

ColGalT identification. Most glycosylation pathways have been characterized at the molecular level over the past decades. However, the genes encoding the glycosyltransferases involved in the glycosylation of Hyl in collagen have remained unknown up to now. Here, we have applied a cloning strategy based on the enrichment of proteins by affinity chromatography, peptide sequencing and heterologous expression of isolated candidate proteins. The enrichment procedure followed the method of Myllyla *et al.* (31) using immobilized denatured collagen type I to capture the ColGalT. Homogenates of 10-day-old chicken embryos were used as source of ColGalT enzyme as applied previously (1, 23). To identify potential ColGalT enzymes among the proteins enriched by the affinity chromatography, we selected proteins sharing sequence homology with known glycosyltransferases. We also narrowed down on proteins containing ER-localization motifs considering the cellular localization of collagen glycosylation and on proteins containing N-glycosylation sites since it was shown that the ColGalT activity could by enriched by concanavalin-A lectin chromatography (29).

One of the candidate proteins identified by tandem-MS peptide sequencing was the putative glycosyltransferase GLT25D2 (Fig. 1) (for further MS hits see supplementary Table A1). GLT25D2 is a type II transmembrane protein of 626 amino acids including four N-glycosylation sites and the ER-retention signal RDEL at the C-terminus. No enzymatic activity was attributed to GLT25D2 but database annotations pointed to sequence homology with

bacterial enzymes involved in LPS biosynthesis. Similar proteins to the chicken GLT25D2 could be deduced from all metazoan genomes. In the human genome, GLT25D2 was found to be strongly similar to two proteins, namely GLT25D1 and CEECAM1. The three proteins contained N-glycosylation sites and the ER retrieval signal RDEL at the C-terminus and shared above 50 % sequence identity (Fig. 2).

ColGalT activity. The putative ColGalT activity of GLT25D1, GLT25D2 and CEECAM1 was assayed by expressing the three proteins as recombinant baculovirus in Sf9 insect cells. Five types of collagen were tested as possible acceptor substrates. Because native collagen is readily glycosylated to varying extents, we also included deglycosylated collagen preparations in the assays. GLT25D1 and GLT25D2 showed a strong ColGalT activity on all deglycosylated collagen acceptors tested, whereas CEECAM1 did not show any activity (Table 1). As expected, the ColGalT activity of GLT25D1 and GLT25D2 was lower when using native collagen acceptors. Noteworthy, collagen type IV and collagen type V in the native form were hardly galactosylated by GLT25D1 and GLT25D2, suggesting that most Hyl residues were already glycosylated. In addition to true collagens, GLT25D1 and GLT25D2, but not CEECAM1, were able to transfer Gal to the serum protein MBL (Fig. 3), which contains four Hyl sites in its collagen domain, showing that ColGalT activity was not limited to large collagen acceptors.

We also tested the ColGalT activity of the human LH3 enzyme, which had been reported to catalyze three reactions on collagen, namely the hydroxyla-

tion of Lys, plus the β(1-O)galactosylation and α(1-2)glucosylation of Hyl (11, 46). Surprisingly, we could not detect any significant ColGalT activity for LH3 under our assay conditions using native collagen type I as acceptor. However, as described previously (11, 46), we did measure a low collagen glucosyltransferase activity for LH3, whereas GLT25D1, GLT25D2 and CEECAM1 failed to show any collagen glucosyltransferase activity (Fig. 4). The expression of CEECAM1 and the other recombinantly expressed enzymes was proven by tryptic digest and LC-MS analysis of Coomassie SDS-PAGE excised protein bands (Fig. 4C, LC-MS hits see supplementary figure A2). Reasons why we could not show the ColGalT activity and only a very low ColGlcT activity of LH3 could be the different collagen acceptor substrate preparations, or differences in the expression system. The apparent Km of GLT25D1 and GLT25D2 was determined for the native collagen type I acceptor substrate and for UDP-Gal, since these values had been reported previously for the semi-purified ColGalT activity (22). The Km of GLT25D1 and GLT25D2 for the collagen type I acceptor was 13.6 g/l and 9.8 g/l, respectively, whereas Myllyla *et al.* reported a Km of 150 g/l for the partially purified chicken ColGalT enzyme. The Km values for UDP-Gal were 18.77 µM for GLT25D1 and 33.53 µM for GLT25D2. These values are comparable with Myllyla *et al*, who determined a Km for UDP-Gal of 30 µM for the partially purified chicken ColGalT (Fig. 5) (22).

ColGalT reaction products. The products of the GLT25D1 and GLT25D2 mediated ColGalT reactions were further analyzed to confirm the transfer of Gal

to Hyl residues on collagen. The reaction products were hydrolyzed in 4 M NaOH to yield single amino acids. After derivatization with Fmoc and separation by reverse phase HPLC, the amino acid profiles obtained from collagen type I, collagen type II and from the GLT25D1-, GLT25D2-reacted collagen acceptors were compared to a profile of authentic amino acid standards. The amount of GHyl and GGHyl was higher in collagen type II than in collagen type I as measured by the ratio GHyl/Hyl and GGHyl/Hyl, respectively (Fig. 6). This finding was in agreement with the values reported in the literature (15). The analysis of additional types of collagen, such as collagen type IV and sponge collagen confirmed the variable extent of Hyl glycosylation across collagens (data not shown). The amino acid profiles obtained after GLT25D1 and GLT25D2 reactions in the presence of UDP-[^{14}C]Gal were further analyzed by β-counting (Fig. 6, lower panel). The [^{14}C]Gal signal co-migrated with the GHyl standard.

ColGalT gene expression. The expression of *GLT25D1*, *GLT25D2* and *CEECAM1* in human tissues was determined by Northern Dot blotting using RNA arrays featuring 75 different human tissues and cell types and by a Northern Blot with mRNAs of 10 different organs (Fig. 7A). The *GLT25D1* gene was widely expressed in fetal and adult human tissues (Fig. 7B). By contrast, the *GLT25D2* gene was expressed at much lower levels and only in the nervous system (Fig. 7B). The *CEECAM1* gene, which could not be related to a ColGalT activity, showed a widespread expression across tissues, whereas *CEECAM1* was highly expressed in secretory tissues like salivary

glands, pancreas and placenta and in the nervous system (Fig. 7B). Additionally, the *GLT25D1* and *CEECAM1* genes, but not *GLT25D2*, were also expressed in various carcinoma cell lines (supplementary Fig. A1). This survey of ColGalT gene expression suggested that GLT25D1 represents the main source of ColGalT activity in the human organism, while GLT25D2 appears to be specialized to few cell types and possibly to few collagen acceptors.

Discussion

The identification of the GLT25D1 and GLT25D2 ColGalT enzymes and of the similar, yet inactive, relative CEECAM1 protein raises the question of a possible restricted specificity towards collagen acceptor substrates. Previous work has shown that ColGalTs recognizes collagen peptides of at least 500-600 Da and that Hyl alone is not a suitable acceptor (22, 30). The different activity levels of recombinant GLT25D1 and GLT25D2 enzymes towards the various types of collagen tested support the idea of a differential substrate recognition. The nature of the substrate recognition may be complex and may include glycosylated residues as part of the motifs recognized. In fact, GLT25D2 was more active towards native collagen type I and type II than towards the deglycosylated forms of these proteins (Table 1), which indicates that glycan chains somehow affected the recognition of the acceptor substrate by GLT25D2. Along the same line, the apparent inactivity of CEECAM1 may be due to stringent structural requirements regarding the recognition of collagen peptides. A similarly complex mechanism of substrate recognition has been described for core glycosyltransferases acting on mucin proteins. Some members of the polypeptide N-acetylgalactosaminyltransferase family recognize peptide acceptors including serine or threonine residues near residues that were previously glycosylated by other N-acetylgalactosaminyltransferases (3). ColGalT assays with synthetic peptides including Hyl and GHyl at various positions will certainly answer the question of the acceptor substrate recognition.

Alternatively, it is possible that CEECAM1 represents a ColGalT acting on a limited set of substrates. The screening of additional proteins including collagen domains like adiponectin, the acetylcholine esterase complex COLQ, and the complement protein C1q might confirm this eventuality. Finally, CEECAM1 may have lost any enzymatic activity over the course of evolution. In fact, CEECAM1 was first described as an adhesion protein (40), which might function as a carbohydrate-binding protein at the cell surface. However, the presence of the C-terminal RDEL motif would suggest that CEECAM1 is maintained in the ER.

The lysylhydroxylase LH3 protein has been previously reported to possess three enzymatic activities, namely a lysylhydroxylase, a ColGalT and a collagen glucosyltransferase activity (46). The glycosyltransferase activities attributed to LH3 were very low, casting doubt on their biological significance (26). By comparison, the strong ColGalT activity described here for GLT25D1 and GLT25D2 implies that these proteins indeed represent true ColGalT enzymes. The dual glycosyltransferase activity of LH3 certainly requires closer attention as it is expected that the catalysis of both β(1-O) and α(1-2) linkages would require distinct domains responsible for the retaining α(1-2) and inverting β(1-O) glycosyltransferase activities (25).

The *GLT25D1* and *GLT25D2* genes are found on human chromosome 19p13 and chromosome 1q25, respectively. The identification of these two ColGalT genes opens new ways to investigate the biological significance of collagen glycosylation. Knockdown experiments of the GLT25D1 and GLT25D2 or-

tholog in *C. elegans* showed different phenotype aberrations like abnormal locomotion, slow growth and deformed mating organs (14, 35). Knockdown experiments of the LH ortholog showed embryonic lethality (20, 24). These findings in *C. elegans* point out the importance of correct collagen glycosylation in development. In humans, ColGalTs are involved in the pathogenesis of connective tissue disorders linked to chromosomes 19p13 and 1q25, such as psoriasis (18) and epidermolysis bullosa (8, 17). The identification of these ColGalTs enables now further investigations in the relationship between these enzymes and connective tissue disorders.

Acknowledgement

We thank Prof. Dr. Beat Steinmann and Dr. Cecilia Giunta (University Children's Hospital Zurich, Switzerland) for helpful discussions, Dr. B. Roschitzki (Functional Genomic Center Zurich, FGCZ, Switzerland) and the FGCZ for their support and Prof. Dr. R. Tenni (University of Pavia, Italy) for the G-Hyl and GG-Hyl amino acid standards. This work was supported by Telethon Action Suisse, the Wolfermann-Nägeli Foundation and by a grant from the Swiss National Science Foundation (PP00A-106756).

Appendix

Supplementary Table A1: MS mascot search result of ColGalT identification

Protein Hits	Description
UPI0000E23134	Cluster: PREDICTED: similar to keratin 1; n=1; Pan troglodytes|Rep: PREDICTED: similar to keratin 1 - Pan troglodytes
UPI0000E2470B	Cluster: PREDICTED: similar to Keratin, type I cytoskeletal 14 (Cytokeratin-14) (CK-14) (Keratin-14) (K14); n=1; Pan troglodytes|Rep: PREDICTED: similar to Keratin, type I cytoskeletal 14 (Cytokeratin-14) (CK-14) (Keratin-14) (K14) - Pan troglodytes
ZZ_FGCZCont0112	sp|ALBU_BOVIN|
ZZ_FGCZCont0184	gi|229552|prf|| 754920A albumin [Bos primigenius taurus]
ZZ_FGCZCont0156	gi|136425|sp|P00760| TRYP_BOVIN TRYPSINOGEN.
UPI0000E24772	Cluster: PREDICTED: keratin 25D isoform 4; n=1; Pan troglodytes|Rep: PREDICTED: keratin 25D isoform 4 - Pan troglodytes
A5A6P6	Cluster: Keratin 2A; n=1; Pan troglodytes verus|Rep: Keratin 2A - Pan troglodytes verus
UPI000155E6FC	Cluster: PREDICTED: similar to Keratin 6A; n=1; Equus caballus|Rep: PREDICTED: similar to Keratin 6A - Equus caballus
UPI0000E203B1	Cluster: PREDICTED: albumin; n=1; Pan troglodytes|Rep: PREDICTED: albumin - Pan troglodytes
UPI00003690AA	Cluster: PREDICTED: lipocalin 1; n=2; Homo/Pan/Gorilla group|Rep: PREDICTED: lipocalin 1 - Pan troglodytes
UPI0000EB2272	Cluster: Keratin, type I cytoskeletal 10 (Cytokeratin-10) (CK-10) (Keratin-10) (K10) (Epithelial keratin-10).; n=1; Canis lupus familiaris|Rep: Keratin, type I cytoskeletal 10 (Cytokeratin-10) (CK-10) (Keratin-10) (K10) (Epithelial keratin-10). - Ca
UPI0000ECB29B	Cluster: glycosyltransferase 25 domain containing 2; n=1; Gallus gallus|Rep: glycosyltransferase 25 domain containing 2 - Gallus gallus
A5PMF6	Cluster: Glycosyltransferase 25 family member 1 precursor; n=1; Danio rerio|Rep: Glycosyltransferase 25 family member 1 precursor - Danio rerio (Zebrafish) (Brachydanio rerio)
UPI000069E8D4	Cluster: glycosyltransferase 25 domain containing 2; n=1; Xenopus tropicalis|Rep: glycosyltransferase 25 domain containing 2 - Xenopus tropicalis
UPI0000D9CC7A	Cluster: PREDICTED: keratin 5 isoform 8; n=1; Macaca mulatta|Rep: PREDICTED: keratin 5 isoform 8 - Macaca mulatta
Q0MRP5	Cluster: Lysozyme C; n=2; Bovidae|Rep: Lysozyme C - Bos indicus x Bos taurus (hybrid cattle)
P08515	Cluster: Glutathione S-transferase class-mu 26 kDa isozyme; n=2; Schistosoma japonicum|Rep: Glutathione S-transferase class-mu 26 kDa isozyme - Schistosoma japonicum (Blood fluke)
Q5XKQ4	Cluster: Alpha-2-glycoprotein 1, zinc-binding; n=1; Homo sapiens|Rep: Alpha-2-glycoprotein 1, zinc-binding - Homo sapiens (Human)
A1SFN4	Cluster: Putative uncharacterized protein; n=1; Nocardioides sp. JS614|Rep: Putative uncharacterized protein - Nocardioides sp. (strain BAA-499 / JS614)
A0A9F3	Cluster: Prolactin-induced protein; n=1; Homo sapiens|Rep:

	Prolactin-induced protein - Homo sapiens (Human)
UPI0000DC1FE0	Cluster: RGD1311906 protein; n=1; Rattus norvegicus|Rep: RGD1311906 protein - Rattus norvegicus
Q5BZZ6	Cluster: SJCHGC01457 protein; n=1; Schistosoma japonicum|Rep: SJCHGC01457 protein - Schistosoma japonicum (Blood fluke)
UPI0000F338C8	Cluster: Keratin, type II cytoskeletal 5 (Cytokeratin-5) (CK-5) (Keratin-5) (K5).; n=1; Bos taurus|Rep: Keratin, type II cytoskeletal 5 (Cytokeratin-5) (CK-5) (Keratin-5) (K5). - Bos Taurus
UPI0000E1FC56	Cluster: PREDICTED: lactotransferrin isoform 5; n=1; Pan troglodytes|Rep: PREDICTED: lactotransferrin isoform 5 - Pan troglodytes
A6UD50	Cluster: Sarcosine oxidase, alpha subunit family; n=1; Sinorhizobium medicae WSM419|Rep: Sarcosine oxidase, alpha subunit family - Sinorhizobium medicae (strain WSM419) (Ensifer medicae)
A3TU89	Cluster: Ferric anguibactin transport system permease protein fatc; n=1; Oceanicola batsensis HTCC2597|Rep: Ferric anguibactin transport system permease protein fatc - Oceanicola batsensis HTCC2597
Q55FK8	Cluster: Putative uncharacterized protein; n=1; Dictyostelium discoideum|Rep: Putative uncharacterized protein - Dictyostelium discoideum (Slime mold)
Q0YHZ7	Cluster: Geobacter sulfurreducens, CxxxxCH...CXXCH motif; n=1; Geobacter sp. FRC-32|Rep: Geobacter sulfurreducens, CxxxxCH...CXXCH motif - Geobacter sp. FRC-32
A0TMF4	Cluster: Putative uncharacterized protein; n=1; Burkholderia ambifaria MC40-6|Rep: Putative uncharacterized protein - Burkholderia ambifaria MC40-6

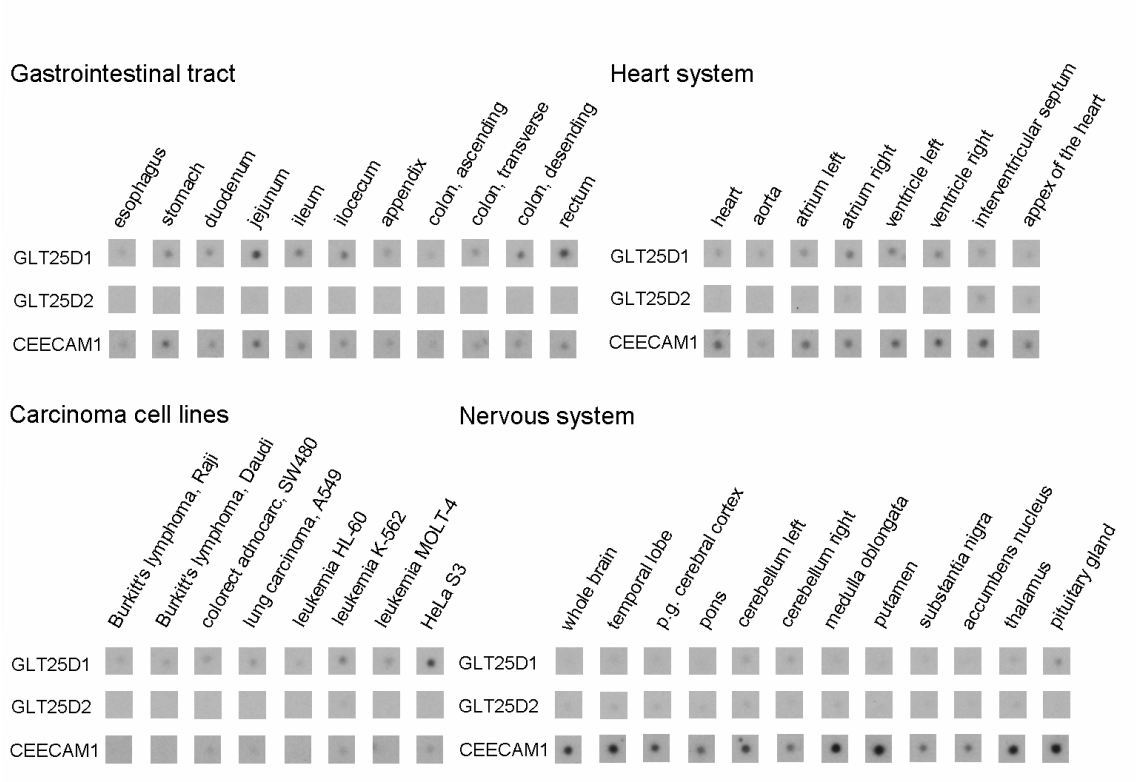
Supplementary Table A2: HPLC gradient

Time [min]	Solvent A [%]	Solvent B [%]	Solvent C [%]
0.0	75	0	25
74.2	64	0	36
78.7	64	0	36
78.9	0	65	35
104.0	0	65	35
129.3	0	25	75
135.3	0	25	75
153.1	75	0	25
188.9	75	0	25

Supplementary Figure A1: Tissue Northern blot. The mRNA expression patterns of GLT25D1, GLT25D2 and CEECAM1 were analyzed in 75 human tissues and cell lines. The 39 tissues and cell lines not presented in Fig. 7 are shown here.

Supplementary Figure A2: MS mascot search result. Proteins expressed in Sf9 cells using the baculovirus system were separated on a SDS-PAGE gel. Protein bands were cut from the gel, digested with trypsin and analyzed by LC-MS. The MS mascot search result of recombinantly expressed GLT25D1, GLT25D2, CEECAM1 and LH3 is shown.

Supplementary Figure A1



Supplementary Figure A2

References

1. **Anttinen, H.** 1976. Stimulation of collagen galactosyltransferase and glucosyltransferase activities by lysophosphatidylcholine. Biochem J **160:**29-35.
2. **Bank, R. A., B. Beekman, R. Tenni, and J. M. TeKoppele.** 1997. Pre-column derivatisation method for the measurement of glycosylated hydroxylysines of collagenous proteins. J Chromatogr B Biomed Sci Appl **703:**267-272.
3. **Bennett, E. P., H. Hassan, U. Mandel, E. Mirgorodskaya, P. Roepstorff, J. Burchell, J. Taylor-Papadimitriou, M. A. Hollingsworth, G. Merkx, A. G. van Kessel, H. Eiberg, R. Steffensen, and H. Clausen.** 1998. Cloning of a human UDP-N-acetyl-alpha-D-Galactosamine:polypeptide N-acetylgalactosaminyltransferase that complements other GalNAc-transferases in complete O-glycosylation of the MUC1 tandem repeat. J Biol Chem **273:**30472-30481.
4. **Bon, S., A. Ayon, J. Leroy, and J. Massoulie.** 2003. Trimerization domain of the collagen tail of acetylcholinesterase. Neurochem Res **28:**523-535.
5. **Cabral, W. A., W. Chang, A. M. Barnes, M. Weis, M. A. Scott, S. Leikin, E. Makareeva, N. V. Kuznetsova, K. N. Rosenbaum, C. J. Tifft, D. I. Bulas, C. Kozma, P. A. Smith, D. R. Eyre, and J. C. Marini.** 2007. Prolyl 3-hydroxylase 1 deficiency causes a recessive metabolic bone disorder resembling lethal/severe osteogenesis imperfecta. Nat Genet **39:**359-365.
6. **Colley, K. J., and J. U. Baenziger.** 1987. Post-translational modifications of the core-specific lectin. Relationship to assembly, ligand binding, and secretion. J Biol Chem **262:**10296-10303.
7. **Edge, A. S., C. R. Faltynek, L. Hof, L. E. Reichert, Jr., and P. Weber.** 1981. Deglycosylation of glycoproteins by trifluoromethanesulfonic acid. Anal Biochem **118:**131-137.

8. **Gay, S., J. D. Fine, and J. S. Storer.** 1988. Autoantibodies to extracellular collagen matrix components in epidermolysis bullosa and other bullous diseases. Arch Dermatol Res **280:**333-337.

9. **Grassmann W, S. H.** 1953. Biochem Z **277:**320-328.

10. **Hakansson, K., and K. B. Reid.** 2000. Collectin structure: a review. Protein Sci **9:**1607-1617.

11. **Heikkinen, J., M. Risteli, C. Wang, J. Latvala, M. Rossi, M. Valtavaara, and R. Myllyla.** 2000. Lysyl hydroxylase 3 is a multifunctional protein possessing collagen glucosyltransferase activity. J Biol Chem **275:**36158-36163.

12. **Hennet, T., A. Dinter, P. Kuhnert, T. S. Mattu, P. M. Rudd, and E. G. Berger.** 1998. Genomic cloning and expression of three murine UDP-galactose: beta-N-acetylglucosamine beta1,3-galactosyltransferase genes. J Biol Chem **273:**58-65.

13. **Ihme, A., L. Risteli, T. Krieg, J. Risteli, U. Feldmann, K. Kruse, and P. K. Muller.** 1983. Biochemical characterization of variants of the Ehlers-Danlos syndrome type VI. Eur J Clin Invest **13:**357-362.

14. **Kamath, R. S., A. G. Fraser, Y. Dong, G. Poulin, R. Durbin, M. Gotta, A. Kanapin, N. Le Bot, S. Moreno, M. Sohrmann, D. P. Welchman, P. Zipperlen, and J. Ahringer.** 2003. Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. Nature **421:**231-237.

15. **Kivirikko, K. I., and R. Myllyla.** 1979. Collagen glycosyltransferases. Int Rev Connect Tissue Res **8:**23-72.

16. **Kukkola, L., R. Hieta, K. I. Kivirikko, and J. Myllyharju.** 2003. Identification and characterization of a third human, rat, and mouse collagen prolyl 4-hydroxylase isoenzyme. J Biol Chem **278:**47685-47693.

17. **Lane, E. B., and W. H. McLean.** 2004. Keratins and skin disorders. J Pathol **204:**355-366.

18. **Lee, Y. A., F. Ruschendorf, C. Windemuth, M. Schmitt-Egenolf, A. Stadelmann, G. Nurnberg, M. Stander, T. F. Wienker, A. Reis, and**

H. Traupe. 2000. Genomewide scan in german families reveals evidence for a novel psoriasis-susceptibility locus on chromosome 19p13. Am J Hum Genet **67:**1020-1024.

19. **Mizuno, K., T. Hayashi, and H. P. Bachinger.** 2003. Hydroxylation-induced stabilization of the collagen triple helix. Further characterization of peptides with 4(R)-hydroxyproline in the Xaa position. J Biol Chem **278:**32373-32379.
20. **Myllyharju, J., and K. I. Kivirikko.** 2004. Collagens, modifying enzymes and their mutations in humans, flies and worms. Trends Genet **20:**33-43.
21. **Myllyla, R., H. Anttinen, L. Risteli, and K. I. Kivirikko.** 1977. Isolation of collagen glucosyltransferase as a homogeneous protein from chick embryos. Biochim Biophys Acta **480:**113-121.
22. **Myllyla, R., L. Risteli, and K. I. Kivirikko.** 1975. Assay of collagen-galactosyltransferase and collagen-glucosyltransferase activities and preliminary characterization of enzymic reactions with transferases from chick-embryo cartilage. Eur J Biochem **52:**401-410.
23. **Myllyla, R., L. Risteli, and K. I. Kivirikko.** 1976. Collagen glucosyltransferase. Partial purification and characterization of the enzyme from whole chick embryos and chick-embryo cartilage. Eur J Biochem **61:**59-67.
24. **Norman, K. R., and D. G. Moerman.** 2000. The let-268 locus of Caenorhabditis elegans encodes a procollagen lysyl hydroxylase that is essential for type IV collagen secretion. Dev Biol **227:**690-705.
25. **Qasba, P. K., B. Ramakrishnan, and E. Boeggeman.** 2005. Substrate-induced conformational changes in glycosyltransferases. Trends Biochem Sci **30:**53.
26. **Rautavuoma, K., K. Takaluoma, K. Passoja, A. Pirskanen, A. P. Kvist, K. I. Kivirikko, and J. Myllyharju.** 2002. Characterization of three fragments that constitute the monomers of the human lysyl

hydroxylase isoenzymes 1-3. The 30-kDa N-terminal fragment is not required for lysyl hydroxylase activity. J Biol Chem **277:**23084-23091.

27. **Reiser, K., R. J. McCormick, and R. B. Rucker.** 1992. Enzymatic and nonenzymatic cross-linking of collagen and elastin. Faseb J **6:**2439-2449.
28. **Richards, A. A., T. Stephens, H. K. Charlton, A. Jones, G. A. Macdonald, J. B. Prins, and J. P. Whitehead.** 2006. Adiponectin multimerization is dependent on conserved lysines in the collagenous domain: evidence for regulation of multimerization by alterations in posttranslational modifications. Mol Endocrinol **20:**1673-1687.
29. **Risteli, L.** 1978. Further characterization of collagen galactosyltransferase from chick embryos. Biochem J **169:**189-196.
30. **Risteli, L., R. Myllya, and K. I. Kivirikko.** 1976. Partial purification and characterization of collagen galactosyltransferase from chick embryos. Biochem J **155:**145-153.
31. **Risteli, L., R. Myllyla, and K. I. Kivirikko.** 1976. Affinity chromatography of collagen glycosyltransferases on collagen linked to agarose. Eur J Biochem **67:**197-202.
32. **Risteli, M., O. Niemitalo, H. Lankinen, A. H. Juffer, and R. Myllyla.** 2004. Characterization of collagenous peptides bound to lysyl hydroxylase isoforms. J Biol Chem **279:**37535-37543.
33. **Shevchenko, A., H. Tomas, J. Havlis, J. V. Olsen, and M. Mann.** 2006. In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat Protoc **1:**2856-2860.
34. **Shinkai, H., and K. Yonemasu.** 1979. Hydroxylysine-linked glycosides of human complement subcomponent C1q and various collagens. Biochem J **177:**847-852.
35. **Simmer, F., C. Moorman, A. M. van der Linden, E. Kuijk, P. V. van den Berghe, R. S. Kamath, A. G. Fraser, J. Ahringer, and R. H. Plasterk.** 2003. Genome-wide RNAi of C. elegans using the

hypersensitive rrf-3 strain reveals novel gene functions. PLoS Biol **1:**77-84.

36. **Soderhall, C., I. Marenholz, T. Kerscher, F. Ruschendorf, J. Esparza-Gordillo, M. Worm, C. Gruber, G. Mayr, M. Albrecht, K. Rohde, H. Schulz, U. Wahn, N. Hubner, and Y. A. Lee.** 2007. Variants in a novel epidermal collagen gene (COL29A1) are associated with atopic dermatitis. PLoS Biol **5:**19521961.

37. **Sojar, H. T., and O. P. Bahl.** 1987. Chemical deglycosylation of glycoproteins. Methods Enzymol **138:**341-350.

38. **Spiro, R. G.** 1969. Characterization and quantitative determination of the hydroxylysine-linked carbohydrate units of several collagens. J Biol Chem **244:**602-612.

39. **Spiro, R. G.** 1967. The structure of the disaccharide unit of the renal glomerular basement membrane. J Biol Chem **242:**4813-4823.

40. **Starzyk, R. M., C. Rosenow, J. Frye, M. Leismann, E. Rodzinski, S. Putney, and E. I. Tuomanen.** 2000. Cerebral cell adhesion molecule: a novel leukocyte adhesion determinant on blood-brain barrier capillary endothelium. J Infect Dis **181:**181-187.

41. **Tenni, R., D. Rimoldi, G. Zanaboni, G. Cetta, and A. A. Castellani.** 1984. Hydroxylysine glycosides: preparation and analysis by reverse phase high performance liquid chromatography. Ital J Biochem **33:**117-127.

42. **Valtavaara, M., C. Szpirer, J. Szpirer, and R. Myllyla.** 1998. Primary structure, tissue distribution, and chromosomal localization of a novel isoform of lysyl hydroxylase (lysyl hydroxylase 3). J Biol Chem **273:**12881-12886.

43. **van der Slot, A. J., A. M. Zuurmond, A. F. Bardoel, C. Wijmenga, H. E. Pruijs, D. O. Sillence, J. Brinckmann, D. J. Abraham, C. M. Black, N. Verzijl, J. DeGroot, R. Hanemaaijer, J. M. TeKoppele, T. W. Huizinga, and R. A. Bank.** 2003. Identification of PLOD2 as

telopeptide lysyl hydroxylase, an important enzyme in fibrosis. J Biol Chem **278:**40967-40972.

44. **Veit, G., B. Kobbe, D. R. Keene, M. Paulsson, M. Koch, and R. Wagener.** 2006. Collagen XXVIII, a novel von Willebrand factor A domain-containing protein with many imperfections in the collagenous domain. J Biol Chem **281:**3494-3504.

45. **Vranka, J. A., L. Y. Sakai, and H. P. Bachinger.** 2004. Prolyl 3-hydroxylase 1, enzyme characterization and identification of a novel family of enzymes. J Biol Chem **279:**23615-23621.

46. **Wang, C., H. Luosujarvi, J. Heikkinen, M. Risteli, L. Uitto, and R. Myllyla.** 2002. The third activity for lysyl hydroxylase 3: galactosylation of hydroxylysyl residues in collagens in vitro. Matrix Biol **21:**559-566.

47. **Zhou, D., M. Malissard, E. G. Berger, and T. Hennet.** 2000. Secretion and purification of recombinant beta1-4 galactosyltransferase from insect cells using pFmel-protA, a novel transposition-based baculovirus transfer vector. Arch Biochem Biophys **374:**3-7.

Table 1. ColGalT activity measured in Sf9 cell lysates. Values represent the average ± S.D. of 4 assays.

Collagen	TFMS [a]	Sf9 [b]	GLT25D1	GLT25D2	CEECAM1
			(pmol·min^{-1}·mg prot^{-1})		
no acceptor		1.5 ± 0.6	1.9 ± 0.7	0.9 ± 0.4	0.5 ± 0.1
Type I	-	10.5 ± 1.8	43.7 ± 12.0	57.2 ± 8.5	8.7 ± 2.0
	+	5.0 ± 3.0	61.4 ± 18.2	18.0 ± 5.4	2.7 ± 1.9
Type II	-	6.5 ± 0.8	36.6 ± 3.7	35.1 ± 3.8	3.3 ± 1.3
	+	3.4 ± 0.8	67.0 ± 16.5	22.1 ± 8.8	2.8 ± 0.8
Type III	-	2.9 ± 0.9	31.7 ± 5.8	25.2 ± 6.8	1.8 ± 0.8
	+	11.0 ± 3.6	244.7 ± 25.7	102.1 ± 17.6	6.5 ± 1.7
Type IV	-	1.7 ± 1.4	3.6 ± 2.2	3.6 ± 0.9	1.9 ± 0.7
	+	12.9 ± 4.6	433.2 ± 49.6	164.2 ± 43.3	8.4 ± 2.4
Type V	-	2.3 ± 0.7	2.7 ± 0.3	2.3 ± 0.4	0.7 ± 0.1
	+	2.7 ± 1.6	18.6 ± 3.6	4.7 ± 1.5	2.0 ± 1.1

[a] TFMS-mediated deglycosylation

[b] Sf9 cells infected with an empty baculovirus

Figure Legends

Fig. 1: ColGalT identification by mass spectrometry. Proteins isolated by affinity chromatography were analyzed by LC-MS. (A) Peptide fragment spectra of two peptides identifying GLT25D2, (B) protein sequence of the *Gallus gallus* GLT25D2. The two identifying peptides are shaded in grey, the four potential N-glycosylation sites are underlined and the ER retrieval signal is shown in bold.

Fig. 2: Protein alignment. The three putative human ColGalT enzymes share a high degree of sequence identity (63 % between GLT25D1 and GLT25D2, 50% between GLT25D2 and CEECAM1, 55 % between GLT25D1 and CEECAM1). The proteins include the C-terminal RDEL ER retrieval motif. Black squares represent amino acids identical or similar in all 3 proteins; grey squares represent amino acids identical or similar in 2 of the proteins.

Fig. 3: ColGalT activity towards MBL. ColGalT activity assay was performed as described in the *Experimental procedures* section. Bars indicate the means of four assays. Error bars indicate the SD.

Fig. 4: Time course of baculovirus mediated protein expression in Sf9 cells. (A) ColGalT activity and (B) collagen glucosyltransferase activity measured in cells expressing GLT25D1 (▲), GLT25D2 (✶), CEECAM1 (□), LH3 (●).

Bovine collagen type I was used as acceptor substrate. The activity measured in Sf9 infected with an empty baculovirus is shown in both panels with filled squares (■). Bars indicate the means of four assays. Error bars indicate the SD. (C) Coomassie SDS-PAGE of recombinantly expressed proteins. Arrows indicate the protein bands analyzed by LH-MS.

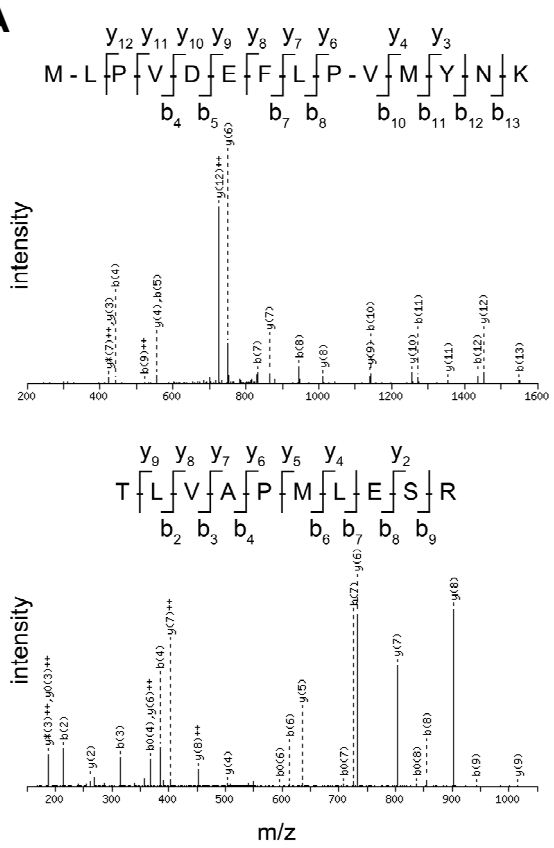
Fig. 5: Michaelis-Menten kinetics. Determination of the apparent Km of GLT25D1 and GLT25D2 for the collagen type I acceptor substrate. (A) Lineweaver Burk blot for GLT25D1 on collagen, with the calculated Michaelis-Menten constant of 13.6 g/l. (C) Lineweaver Burk blot for GLT25D2 on collagen with the calculated Michaelis-Menten constant of 9.8 g/l. (B) Lineweaver Burk blot for GLT25D1 on UDP-Gal with the calculated Michaelis-Menten constant of 18.77 µM . (D) Lineweaver Burk blot for GLT25D2 on UDP-Gal with the calculated Michaelis-Menten constant of 33.53 µM.

Fig. 6: Product identification by RP-HPLC. The first panel represents an amino acid standard containing the standards for G-Hyl and GG-Hyl. The second and the third panels show the amino acid profiles of bovine collagen type I and type II hydrolysates, respectively. The lower two panels show the radioactive trace obtained after reaction of collagen type I with GLT25D1 and GLT25D2. [^{3}H]Val and [^{14}C]Tyr were used as internal amino acid standards. Amino acids are marked in single-letter code. Hyp, hydroxyproline.

Fig. 7: Tissue Northern blot. The mRNA expression patterns of *GLT25D1*, *GLT25D2* and *CEECAM1* were analyzed in 10 human organs (A) (lane 1: brain, 2: placenta, 3: skeletal muscle, 4: heart, 5: kidney, 6: pancreas, 7: liver, 8: lung, 9: spleen, 10: colon) as well as in 75 human tissues and cell lines, whereas a representative collection of 36 tissues is shown in panel (B) (see also supplementary Fig. A1). PBL, peripheral blood leukocytes.

Figure 1

A



B

```
XGSIADHWSEKKSMCISNGRKLILIASRTVSILTIVTHRA    1
SYPGASGWLPLEQMTWHQLPSAVQGCSAILQYTAVGEGPR   41
ILQPSEHGQEGKGSLLQLQQPPAFALFCTLETLIMKYLPL   81
FANGGGWVATDHNADNTTAILREWLKNVQNLYHDVEWRPM  121
EDPQSYPEEMGPKHWPSSRFTHVMKLRQAALRAAREKWSD  161
YVLFLDTDNLLTNPETLNLLIAENKTLVAPMLESRFLYSN  201
FWCGITPQAGGWGYYKRTLDYPLIREWKRTGCFAVPMIHS  241
TFLIDLRKEASTKLMFYPPHQDYTWSFDDIMVFAFSSRQA  281
GIQMFICNREHYGFLPMPLKSHQTLQEETENFVHTLIEAM  321
SKWLCPAVLEPPVSICRHVQLYQSSVYLQVMVGISALFQS  361
IVLPLSPLSMSLFLRSALTRWDLMKALNTSQLKALSIDML  401
PGYRDPYSSRPLTRGEIGCFLSHYYIWKEVVNRGLEKTLV  441
IEDDVRFEHQFKRKLMKLMDDIEQAQLDWELIYIGRKRMQ  481
VQQPEKAVPNVMNLVEADYSYWTLGYAISFQGAQKLIGAE  521
PFSKMLPVDEFLPVMYNKHPVAKYMEYYESRDLKAFSAEP  561
LLVYPTHYTGQPGYLSDTETSTIWDNETVSTDWDRTHSWK  601
SRQQGQIHSEAQNKDALPPQSSLNAPSSRDEL          641
```

Figure 2

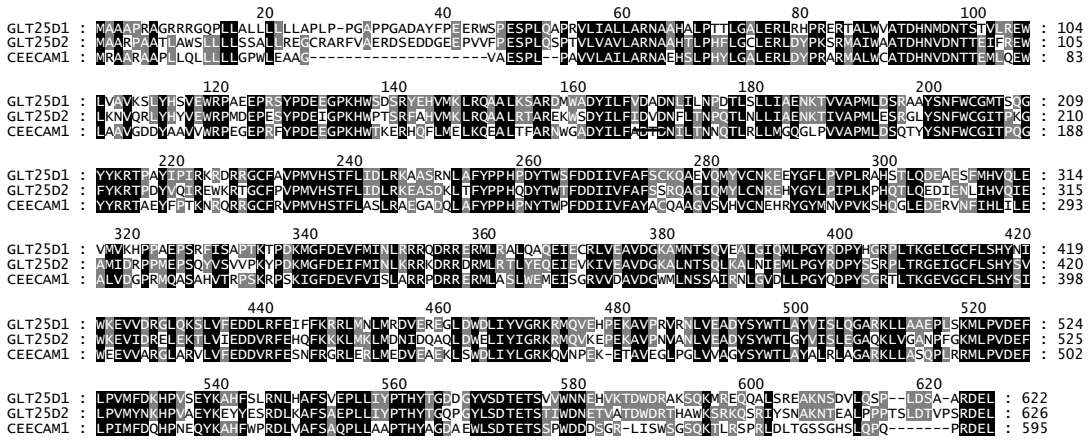


Figure 3

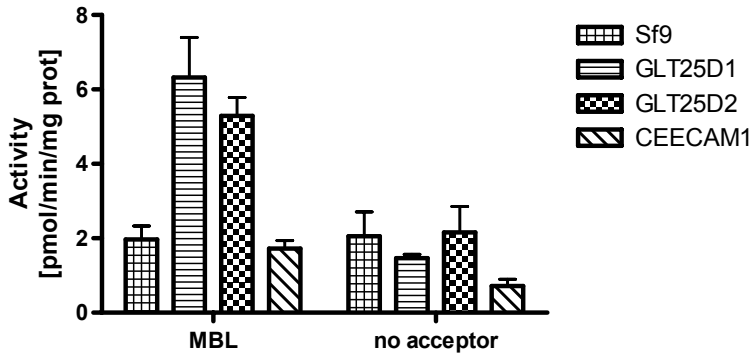

Activity
[pmol/min/mg prot]
8
6
4
2
0
MBL
no acceptor
Sf9
GLT25D1
GLT25D2
CEECAM1

Figure 4

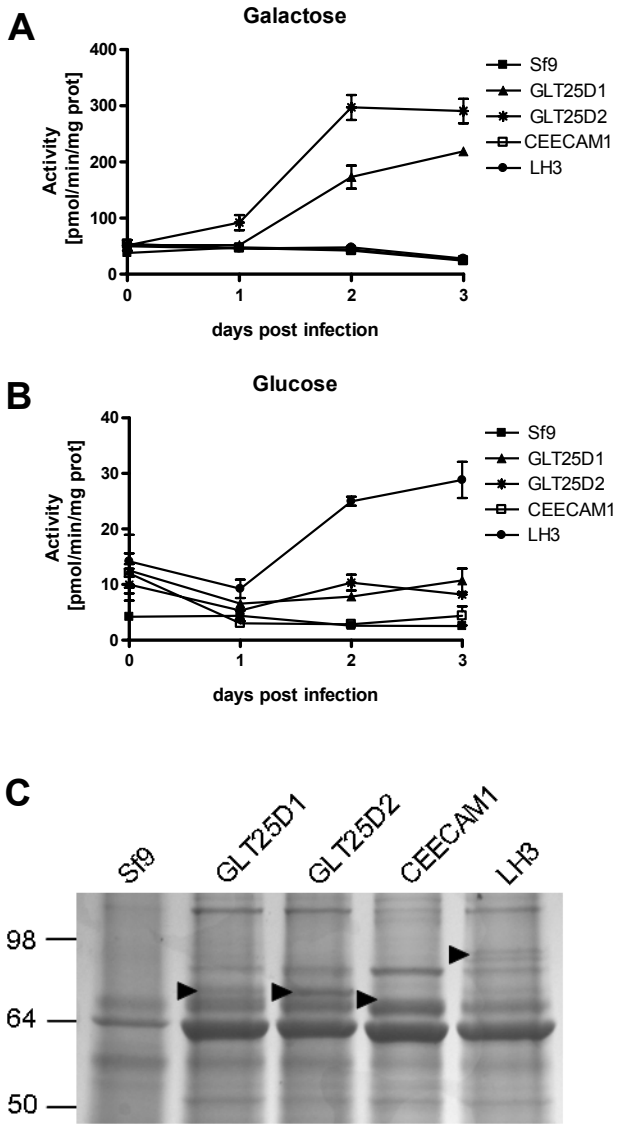

A
Galactose
Activity
[pmol/min/mg prot]
400
300
200
100
0
0
1
2
3
days post infection
Sf9
GLT25D1
GLT25D2
CEECAM1
LH3
B
Glucose
Activity
[pmol/min/mg prot]
40
30
20
10
0
0
1
2
3
days post infection
Sf9
GLT25D1
GLT25D2
CEECAM1
LH3
C
Sf9
GLT25D1
GLT25D2
CEECAM1
LH3
98
64
50

Figure 5

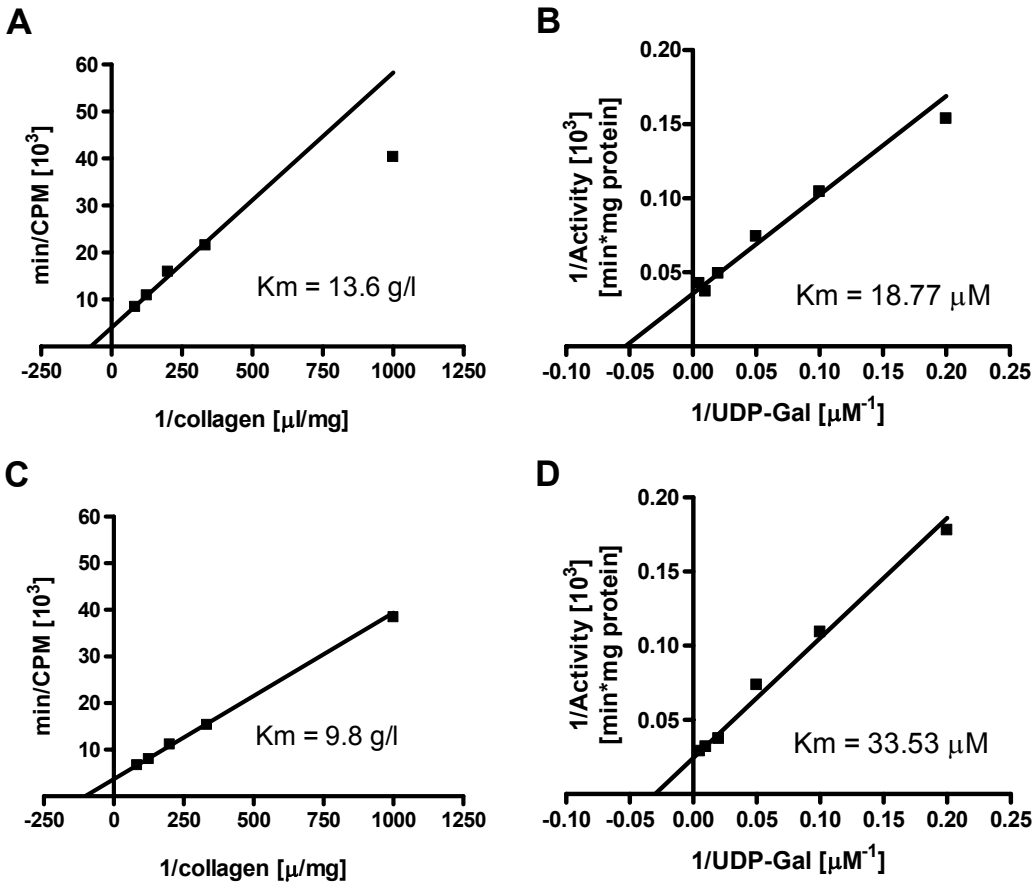

A
min/CPM [10^3]
Km = 13.6 g/l
1/collagen [µl/mg]
B
1/Activity [10^3] [min*mg protein]
Km = 18.77 µM
1/UDP-Gal [µM^-1]
C
min/CPM [10^3]
Km = 9.8 g/l
1/collagen [µ/mg]
D
1/Activity [10^3] [min*mg protein]
Km = 33.53 µM
1/UDP-Gal [µM^-1]

Figure 6

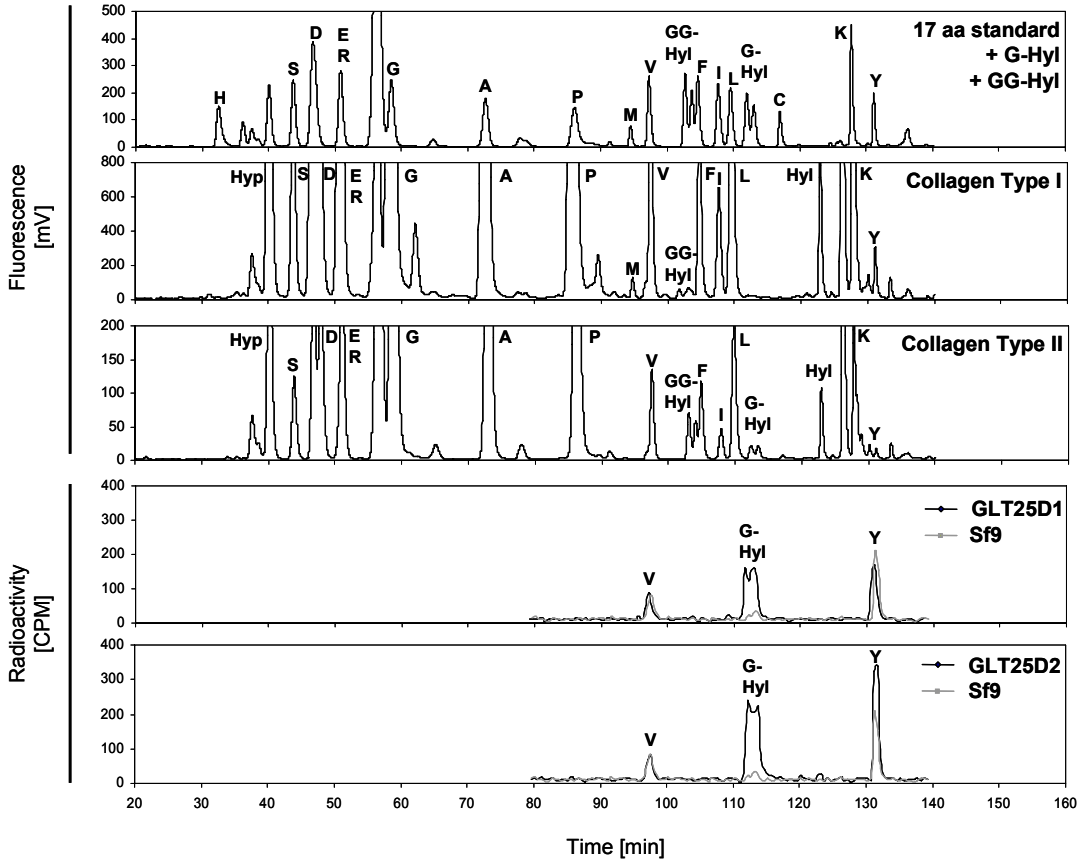
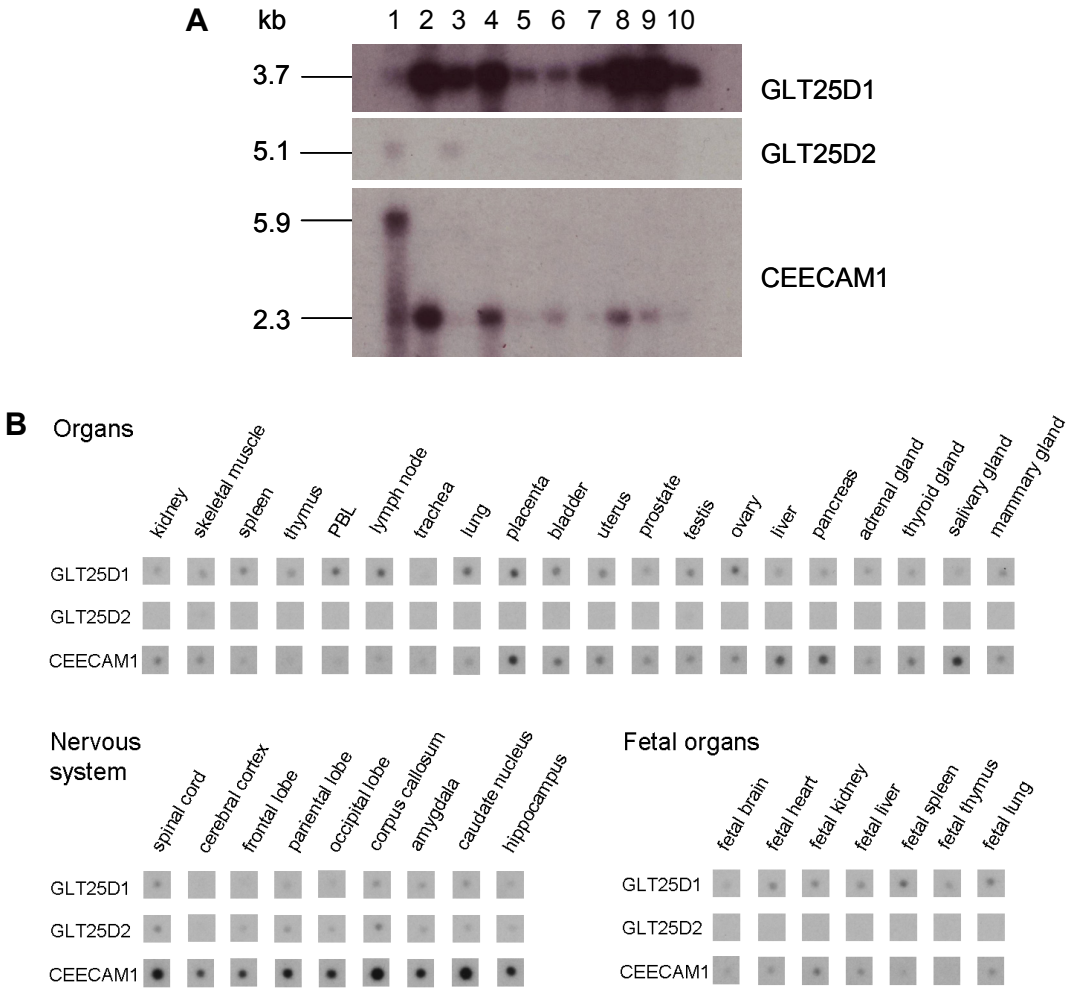

Fluorescence [mV]
17 aa standard
+ G-Hyl
+ GG-Hyl
Collagen Type I
Collagen Type II
Radioactivity [CPM]
GLT25D1
Sf9
GLT25D2
Sf9
Time [min]

Figure 7


A
kb
1 2 3 4 5 6 7 8 9 10
3.7
5.1
5.9
2.3
GLT25D1
GLT25D2
CEECAM1
B
Organs
kidney
skeletal muscle
spleen
thymus
PBL
lymph node
trachea
lung
placenta
bladder
uterus
prostate
testis
ovary
liver
pancreas
adrenal gland
thyroid gland
salivary gland
mammary gland
GLT25D1
GLT25D2
CEECAM1
Nervous system
spinal cord
cerebral cortex
frontal lobe
parietal lobe
occipital lobe
corpus callosum
amygdala
caudate nucleus
hippocampus
Fetal organs
fetal brain
fetal heart
fetal kidney
fetal liver
fetal spleen
fetal thymus
fetal lung

2.1.2 Further Characterization of the Human ColGalTs

2.1.2.1 Synthetic Peptides as Acceptor Substrate

In chapter 2.1.1 we described that GLT25D1 and GLT25D2 are able to glycosylate different acceptor substrates namely collagen type I-V and the non-collagenous protein MBL. In a further step, we have expanded the study of the ColGalT enzymes on additional acceptor substrates. We have investigated synthetic peptides representing sequences containing Hyl, as found in human ADIPOQ, in the α1-chain of human collagen type II (Col2A1) and in the mimivirus collagen L71. These peptides were kindly provided by NeoMPS, PolyPeptide Laboratories Group, France. As shown in the first part of the results section (Chapter 2.1), the GLT25D1 enzyme was able to glycosylate the non-collagenous protein MBL. Using the synthetic peptide acceptors, we could demonstrate that GLT25D1 is also capable of glycosylating the ADIPOQ peptide (Figure 14). Furthermore, the GLT25D1 could glycosylate L71, a collagen of the mimivirus. These results show that the acceptor specificity of the human GLT25D1 enzyme is not restricted to human collagens.

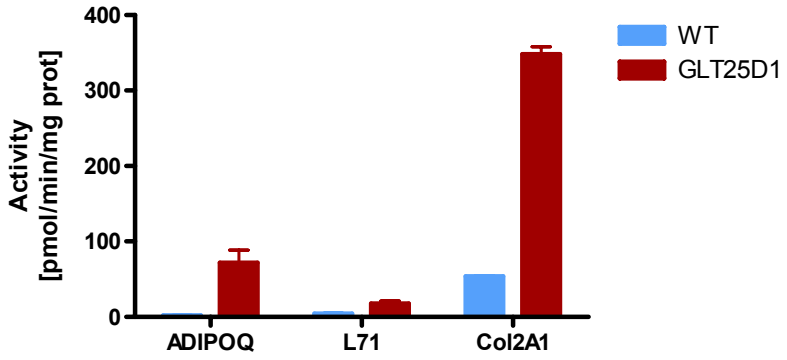


Figure 14: Enzymatic Activity Assay using Synthetic Peptides. Synthetic peptides representing human adiponectin (ADIPOQ), mimiviral L71 and human Col2A1 were used as substrate in the enzymatic activity assay using GLT25D1 as galactosyltransferase. Peptide sequences: ADIPOQ: GDPGLIGP-Hyl-LGDIGETGVP, L71: GDVGD-Hyl-GDVGD-Hyl-GDVGD-Hyl-GD, COL2A1: GPTGVTGP-Hyl-GARGAQGP. Error bars represent standard deviations of n=4.

2.1.3 ColGalT and ColGlcT in Diseases

As reviewed in chapter 1.2.5 many mutations in collagen processing enzymes can lead to a series of diseases. Having in mind that glycans play an important role in many proteins, our hypothesis was that the carbohydrates of

collagen may also play an important role in the function of collagen. In order to investigate this hypothesis, we have been looking for impaired collagen glycosylation in patients showing abnormalities in skin stability.

2.1.3.1 Ehlers-Danlos Syndrome Patients

Since EDS VIb patients show a normal LH activity but share the same phenotype as EDS VIa, we hypothesized that mutations in the collagen glycosyltransferases resulting in an impaired modification of the hydroxylysine could be involved in this disease. As mutations in the collagen glycosyltransferases could be involved in this disease, fibroblasts of EDS VIb patients were investigated (fibroblasts were kindly provided by B. Steinmann, University Children's Hospital Zurich, Switzerland and L. Bonafé, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland). Cell lysates of these patient fibroblasts were used as a source for the collagen glycosyltransferases and assayed on the acceptor collagen type I. Figure 15 shows the enzymatic activity assays using the donor substrates UDP-Gal (panel A) and UDP-Glc (panel B).

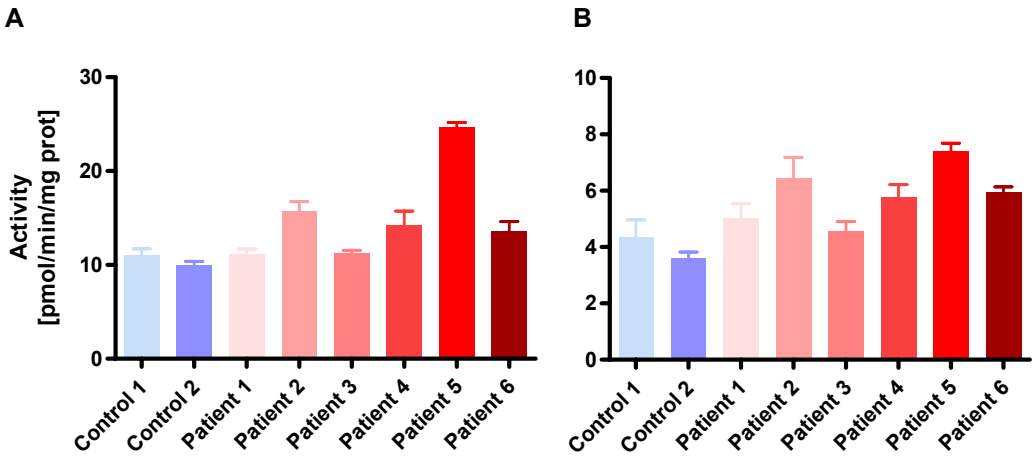


Figure 15: Enzymatic Activity in EDS VIb patients. Six EDS VIb patient fibroblasts were assayed using either UDP-Gal (A) or UDP-Glc (B) as substrate and compared with the two healthy control fibroblasts. Error bars represent standard deviations of n=2-5.

Comparing the enzymatic activities of healthy control fibroblast with the enzymatic activities of the six different EDS VIb patients, no significant reduction of one of the collagen glycosyltransferases was detected as would be expected, if one of these enzymes would be mutated. This indicates that

probably none of the collagen glycosyltransferases are involved in the EDS VIb disease.

2.1.3.2 Brittle Cornea Syndrome Patients

Since BCS patients show a similar phenotype as EDS VI patients and since the cause of the disease is still unknown, fibroblasts of BCS patients were also tested for an altered ColGalT activity (fibroblasts were kindly provided by B. Steinmann, University Children's Hospital Zurich, Switzerland). Figure 16 shows the ColGalT (A) and ColGlcT (B) enzymatic activities measured in BCS patient fibroblasts. Although in patient number 1 a reduction of activity was visible, no loss of ColGalT or ColGlcT activity could be detected as it would be expected if one of the glycosyltransferase activities would be deficient.

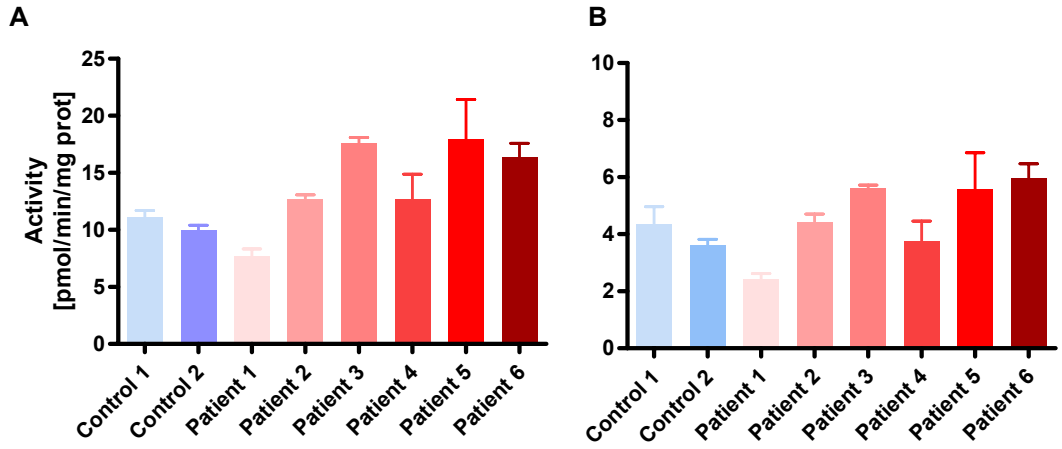


Figure 16: Enzymatic Activity of BCS patients. Six different BCS patient fibroblasts were assayed using either UDP-Gal (A) or UDP-Glc (B) as substrate and compared with the two healthy control fibroblasts. Error bars represent standard deviations of n=3-4.

2.2 Characterization of a Viral ColGlcT

2.2.1 Manuscript 2

Identification and characterization of the collagen glucosyltransferase L230 of *A. polyphaga* mimivirus

Belinda Schegg, Stefan Deuber, Thierry Hennet

Institute of Physiology and Center for Integrative Human Physiology, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich

Own Contribution

Cloning and Baculovirus protein expression

Enzymatic Activity Assays (GT and LH)

HPLC Analysis

Northern Blotting

Identification and characterization of the collagen glucosyltransferase L230 of *A. polyphaga* mimivirus

Running Title: Collagen glucosylation in mimivirus

Belinda Schegg, Stefan A. Deuber, Thierry Hennet§

Institute of Physiology and Center for Integrative Human Physiology, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich

§Corresponding author: Thierry Hennet

Institute of Physiology, University of Zurich,

Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

Tel: +41-1-635-5080, Fax: +41-1-635-6814

E-mail: thennet@access.uzh.ch

Abbreviations: MIMIV, mimivirus; ColGlcT, collagen glucosyltransferase; ColGalT, collagen galactosyltransferase; Hyl, hydroxylysine; GHyl, galactosylated hydroxylysine; GGHyl, glucosyl-galactosylated hydroxylysine

Keywords: *Acanthamoeba polyphaga* mimivirus, collagen, glycosylation, hydroxylation, post-translational modification

Abstract

The *Acanthamoeba polyphaga* mimivirus is the largest virus described to date. Within its 1.2 Mbp linear DNA genome, eight open reading frames with collagen triple helix motifs are encoded. These collagenous proteins are probably involved in the formation of the fibrils surrounding the icosahedral capsid of mimivirus. Animal collagens are hydroxylated and subsequently glycosylated on hydroxylysine. Mimivirus contains genes in its genome, which show similarities to animal lysyl hydroxylase and prolyl hydroxylase genes, which suggests that mimiviral collagens may be also post-translationally modified. Furthermore, genes encoding potential collagen glycosyltransferases could be retrieved from the mimivirus genome. In the present study, we have cloned and characterized the mimiviral *L230* gene. We could show that L230 encodes a collagen glucosyltransferase enzyme, which transfers glucose on hydroxylysines found in collagen motifs. This work represents the first report on a virally encoded collagen glycosyltransferase.

Introduction

The *Acanthamoeba polyphaga* mimivirus (MIMIV) is the largest virus characterized to date. MIMIV belongs to the *Mimiviridiae* family of nucleo-cytoplasmic large DNA viruses. The viral core contains a 1.2 Mbp linear dsDNA genome coding for a total of 911 predicted proteins, which are involved in amino acid transport and metabolism, translation and post-translational modifications (12, 19). The virus has a icosahedral capsid surrounded by fibrils with a total diameter of 650 nm. The 120 nm long external fibrils are unique for MIMIV (19). Raoult (12) and Xiao (23) reported that MIMIV contains eight open reading frames with collagen motifs in its genome, which are most probably found in the fibrils of the virus capsid. The predominant occurrence of lysine in the viral collagen motifs enables the formation of hydroxylysine (Hyl) and glycosylated Hyl. The positive Gram staining of MIMIV particles suggests the presence of carbohydrates on the surface of the virus, thereby supporting the notion that collagen might be glycosylated (12). The process of glycosylation of animal collagens is well known (4). Lysines within the collagen triple helical domain are hydroxylated and the resulting Hyl serve then as acceptors for either the monosaccharide galactose or the disaccharide glucosylgalactose. The MIMIV genome encodes putative enzymes such as lysyl hydroxylase and prolyl hydroxylase, which are involved in the post-translational modifications of collagen (12). This fact suggests that mimiviral collagen are post-translationally hydroxylated and might be glycosylated by virus-encoded enzymes. The present study focuses on the charac-

terization of a MIMIV gene encoding a putative enzyme combining lysyl hydroxylase and collagen glycosyltransferase domains. Collagen or collagen like proteins are not only seen in higher eukaryotes but also in bacteria (13, 14, 24), sponges (7), fungi (2) and even the *Chlorella* virus encodes in its genome a collagen gene (25). Post-translational modifications of collagen, including collagen glycosylation are conserved processes. Glycosylated collagens are not only present in higher animals but are also found in sponges (7).

Materials and Methods

Cloning and protein expression. MIMIV genomic DNA was isolated according to Raoult *et al* (12). *L230* was cloned from viral genomic DNA using the primers 5'- GACCCATGGGATCCATTAGTAGAACTTATGTAAT-3' and 5'- GTCACTAGTTTAATTAACAAAAGACACTAAAATAT-3' (Microsynth). The PCR product was cloned into the baculovirus transfer vector pFastBacI (Invitrogen) by the *Nco*I and *Spe*I restriction sites. The C-terminal deletion construct of L230 (L230 ΔC-term) was amplified with the primers 5'-GTGACTAGGCCTTATTCATACCGTCCCACCATC-3' and 5'-CTAGCTCGAGGCATGAGAGGACCTACAAAA-3' and cloned in pFastBacI with the restriction sites *Stu*I and *Xho*I. The N-terminal deletion construct of L230 (L230 ΔN-term) was amplified with the primers 5'-AAAAAGCAGGCTTAATGATTAGTAGAACTTATG-3' and 5'-AGAAAGCTGGGTATCATTAATTAACAAAAGACAC-3' in the baculovirus transfer vector pDEST8. The human lysyl hydroxylase 1 (*LH1*) cDNA was purchased from the RZPD repository (Berlin, Germany) and cloned by PCR with the primers 5'-AAAAAGCAGGCTTAATGCGGCCCCTGCTGCTA-3' and 5'-AGAAAGCTGGGTATTAGGGATCGACGAAGGA-3' into the vector pDEST8. Recombinant baculoviruses were produced in *S. frugiperda* Sf9 cells as described previously (6).

Preparation of collagen acceptors. Bovine achilles collagen type I (Sigma-Aldrich) was incubated in 0.1 N NaOH for 1 h at 37 °C and brought to pH 6.8 with acetic acid (16). Synthetic peptides including Hyl were designed to

represent sequences derived from human adiponectin ADIPOQ (GDPGLIGP-Hyl-GDIGETGVP), the MIMIV L71 protein (GDVGD-Hyl-GDVGD-Hyl-GDVGD-Hyl-G) and human collagen type II (COL2A1) (GPTGVTGP-Hyl-KGARGAQGP). The peptides were kindly provided by NeoMPS, PolyPeptide Laboratories Group, France and dissolved in H_2O.

Collagen glycosyltransferase assay. Sf9 cells infected with recombinant baculoviruses were harvested three days post infection and lyzed in 1 % Triton-X100 for 10 min on ice. Postnuclear supernatants were used as enzyme source. The collagen acceptors were heat denatured for 10 min at 60 °C and subsequently cooled down to 0 °C. The enzymatic activity assays were performed with 10 µl of the Sf9 postnuclear supernatant in a volume of 100 µl containing 500 µg collagen or 50 µg synthetic peptide acceptor, 60 µM UDP-Galactose (UDP-Gal) or UDP-Glucose (UDP-Glc), 50,000 cpm UDP-[^{14}C]Gal or UDP-[^{14}C]Glc (GE Healthcare), 10 mM $MnCl_2$, 20 mM NaCl, 50 mM MOPS pH 7.4, 1 mM DTT. After an incubation time of 3 h at 37 °C, the reaction was stopped by the addition of 500 µl of ice cold 5 % TCA / 5 % phosphotungstic acid.

Collagen lysyl hydroxylase assay. The lysyl hydroxylase assay was performed according to Kivirikko *et al* (8) with radioactive 2-oxo[1-^{14}C]glutarate (PerkinElmer) and 10 µg synthetic collagen peptide acceptors with the amino acid sequence $(GIK)_4$ (3, 22) and $(GDK)_4$ (Thermo Scientific). The reaction

was incubated for 45 min at 37 °C and the production of [^{14}C]O_2 was determined in a β-counter (Tri-Carb 2900TR, Packard).

Amino acid analysis. The products of collagen glucosyltransferase (ColGlcT) assays were hydrolyzed in 4 M NaOH for 72 h at 105 °C. The resulting single amino acids were derivatized with Fmoc according to Bank *et al.* (1). Separation of Fmoc labelled amino acids was performed on a RP-HPLC (LaChrom Hitachi, Merck) using an ODS Hypersil column, 150 x 3 mm, 3 µm particle size (Thermo Electron Corporation) at 40 °C. The amino acids were separated at a flow rate of 0.2 ml/min using a gradient elution (Table A1) with the solvents 0.5 M citric acid, 5 mM $(CH_3)_4NCl$, pH 2.85 (A); 80 % of 20 mM sodium acetate trihydrate, 5 mM $(CH_3)_4NCl$, pH 4.5, 20 % of methanol (B); 100 % of ACN (C). Radiolabelled [^{3}H]Val and [^{14}C]Tyr (Moravek Biochemicals and Radiochemicals, USA) were used as internal standard. The galactosylated Hyl (GHyl) and galactosyl-glucosylated Hyl (GGHyl) standards were kindly provided by Ruggero Tenni (University of Pavia) (21). The fractions collected from the HPLC runs were counted in a β-counter (Tri-Carb 2900TR, Packard).

Northern blotting. *A. polyphaga* cells were infected with MIMIV for different time periods. 2.5 µg RNA of mock, 4, 8, 16 and 24 hours post infection (hpi) were separated on a 1 % formaldehyde-agarose gel and transferred on a nylon membrane (Hibond-N, GE Healthcare). The *L230, R699* and *R241* probes were amplified by PCR using mimiviral genomic DNA with the

primers 5'-ACTGGGCTTGCGCATATGTC-3', 5'-GGTTGGTGCACATGGA-ATAG-3' giving a product of 515 bp, 5'-TGAAGCTGGAGGTGGACAA-3', 5'-GAGCCGACAGATCATTAG-3' giving a product of 641 bp and 5'-AGGTG-GTGGTGGTGGTATTG-3', 5'-TGAGCGGGAATAGTTGTGGG-3' giving a product of 321 bp, respectively. DNA probes were labelled with α[^{32}P]dCTP (Hartmann Analytic, Germany) by random priming (Stratagene). Membranes were incubated for 2 h at 80 °C and prehybridized for 1 h at 64 °C with QuikHyb® hybridization solution (Stratagene) containing 100 µg/ml ultra pure herring sperm DNA (Invitrogen). Hybridization was performed over-night at 64 °C using 5 x 10^5 cpm per ml hybridization solution of each la-belled probe. Membranes were washed in 0.1 x SSC, 0.1 % SDS up to 60 °C and exposed on BioMax XAR film (Kodak) for 24 h at RT.

Results

ColGlcT identification. Much is known about eukaryotic glycosylation, but only little about virus encoded glycosyltransferases (5, 10, 18). As MIMIV stains positive in the Gram staining (12) and as we suppose that this could arise from the glycosylated collagenous fibrils on the virus capsid surface, we have searched for mimiviral encoded glycosyltransferases. The human collagen galactosyltransferases (ColGalTs) GLT25D1 and GLT25D2 as well as the non-active homologue CEECAM1 (15) were blasted against proteins of other species and aligned using the ClustalW algorithm (Fig. 1). The ColGalTs are conserved within the eukaryotes, but there are also orthologous proteins in a variety of bacteria and even in MIMIV. To identify the mimiviral collagen glycosyltransferases, the MIMIV genome was blasted against the human GLT25D1 and GLT25D2 ColGalTs. In a blast search, the MIMIV proteins R699, R655, R654 and L230 were found as potential mimiviral collagen glycosyltransferases, since these viral proteins contained sequence motifs similar to regions of the human ColGalT enzymes. The putative MIMIV collagen glycosyltransferase L230 showed not only sequence homology to the human ColGalTs GLT25D1 and GLT25D2 (15) in its N-terminal region, but also showed a homology to the human LH1, LH2 and LH3 in its C-terminal domain (Fig. 2). This sequence homology suggests that L230 may possess two activities. It may act as a viral collagen lysyl hydroxylase and also as a viral collagen glycosyltransferase.

ColGlcT activity. The putative collagen glycosyltransferase activity of L230 was assayed using the baculovirus protein expressing system in Sf9 insect cells. Recombinant L230 was tested on bovine collagen type I acceptor for collagen glucosyl- and galactosyltransferase activity. As depicted in figure 3, L230 showed a ColGlcT activity, whereas the enzyme was unable to use UDP-Gal as substrate in spite of its sequence homology to the human Col-GalT enzymes.

Beside bovine collagen type I, three synthetic peptides were tested as acceptor substrates, a peptide derived from the mimiviral collagen L71, from the human Col2A1 and from the collagen domain of human adiponectin (ADIPOQ). The L230 protein was not only able to glucosylate the mimiviral collagen but also the human Col2A1 and the human ADIPOQ peptides (Fig. 4).

ColGlcT product analysis. The product of the enzymatic activity assays using L230 as ColGlcT was further analyzed by reverse phase HPLC to confirm the transfer of Glc to Hyl. The product of the reaction between L230 and the collagen type I acceptor was hydrolyzed in 4 M NaOH and the resulting single amino acids were labelled with Fmoc. The amino acid profile of the reaction product of L230 with UDP-[^{14}C]Glc was further analyzed by β-counting. The profile of radioactive counts (Fig. 5, lower panel) was compared with a 17 amino acid standard containing the standard GHyl and GGHyl and with collagen type II. In animal collagen, Glc is transferred on GHyl. As we could measure a ColGlcT activity for L230, it was expected

from the product analysis that the radioactive counts co-migrate with the GGHyl standard. Surprisingly, [^{14}C]Glc, which was transferred to collagen by L230, co-migrated with GHyl and not GGHyl. Therefore, L230 was shown to encode a ColGlcT that utilizes Hyl as acceptor. Mimiviral collagen glycosylation seems to be different than the mammalian collagen glycosylation as up to now core glucosylation of hydroxylysine in collagen was not reported.

L230 truncated protein constructs. As the L230 protein includes two domains, namely the N-terminal domain homologous to human ColGalTs and the C-terminal domain homologous to human LH, we wanted to explore the enzymatic activities of these domains. To delineate the ColGlcT domain of L230, deletion constructs of either the N-terminal GLT25-like domain or the C-terminal LH-like domain were produced. While the native L230 protein is 103.5 kDa, the truncated proteins L230 ΔC-term and L230 ΔN-term are 70.0 kDa and 67.9 kDa, respectively. These constructs were tested in the enzymatic activity assay using UDP-Glc as the donor substrate. As depicted in figure 6, neither of the two truncated constructs was able to *in vitro* glucosylate collagen, which might be due to incorrect folding or instability of the truncated version of the proteins.

Lysyl hydroxylase activity assay. As the L230 enzyme shows a domain with homology to the LH enzyme, the ability of L230 to act as a LH enzyme was tested. Lysine hydroxylation was tested on synthetic peptide acceptors, ei-

ther the peptide $(GIK)_4$ (3, 22) or the mimiviral peptide $(GDK)_4$. The human lysyl hydroxylase LH1 was used as a positive control in the LH activity assay (20). LH1 was only able to hydroxylate the $(GIK)_4$ peptide but not the mimiviral peptide $(GDK)_4$. By contrast, the L230 enzyme did not show any LH activity on neither of the two peptides tested (Fig. 7).

ColGlcT gene expression. Mimiviral infection of *A. polyphaga* has a typical viral development cycle (9). Newly synthesized virions accumulate in the cytoplasm at 8 hpi. At 24 hpi, most of the infected amoebae are lyzed and large amounts of infectious MIMIV particles are released in the culture medium (19). The onset of RNA expression of mimiviral collagens and of enzymes potentially involved in collagen modification was investigated during infection by Northern Blot analysis. Total RNA was isolated from *A. polyphaga* at 4, 8, 16 and 24 hpi. The expression of the L230 and of the mimiviral collagen glycosyltransferase candidate *R699* genes was compared with the expression of the mimiviral collagen *R241* gene during mimiviral infection. R699 is clearly expressed at 8 hpi, while L230 is only very weakly expressed at this time point. Collagen expression begins 16 hpi and correlates with the expression of the ColGlcT L230 (Fig. 8).

Discussion

Viruses are obligate intracellular parasites and depend on many cellular functions to complete their replication. The limited space for genetic information requires optimized genomes, which are reduced to the essentials. Therefore it is remarkable that the category of large DNA viruses encode many proteins with similar functions to those found in the host cells. MIMIV, which represents the largest virus identified to date, has for example many genes on his own that encode enzymes involved in the translation of proteins or DNA repair (19). Furthermore, the MIMIV genome includes many genes sharing similarity with eukaryotic and prokaryotic glycosylation genes. In this respect, the present identification of the ColGlcT enzyme in MIMIV does not come as a surprise.

In this report, we have shown that the L230 ColGlcT activity is different to the human ColGalT activity. Based on the knowledge from the animal collagen, which is glucosylated only upon galactosylation, we expected that the product of the L230 ColGlcT activity assay would be Glc(α1-2)Gal(β1-O) (17). However, HPLC product analysis of the products from the enzymatic activity assay performed with the L230 ColGlcT revealed that Glc was directly linked to Hyl and not to GHyl (Fig. 5). Further analysis of the carbohydrates will have to confirm this finding.

L230 did not show any LH activity on the tested acceptor peptides, even though L230 is homologous in its C-terminal region to the human LH3 enzyme. The peptide acceptors tested were chosen either from the literature

((GIK)$_4$, (3, 22)) or by considering their high occurrence in MIMIV collagens ((GDK)$_4$). If GDK would represent the recognition sequence of L230 as the lysyl hydroxylase, all the mimiviral collagen would be extensively hydroxylated, which seems unlikely. Other peptide sequences, which are more heterogeneous in their sequence, should be tested as LH acceptor peptides.

To determine whether the GLT25-like and the LH-like domains represent distinct enzymatic entities, the single domains of L230 were produced in Sf9 insect cells and tested for their ColGlcT activity. Neither the N-terminal human GLT25-like domain nor the C-terminal LH-like domain were able to act as ColGlcT. It might be that the truncated proteins are not correctly folded or do aggregate and therefore are not active as enzymes. To circumvent this possibility, a mutagenesis study will allow the delineation of the ColGlcT activity within the L230 protein.

The fibrils on the viral capsid are supposed to have a collagenous origin (9, 12). Analysis of these fibrils of MIMIV would give further information about the mimiviral collagens. Currently it has been reported that the MIMIV genome contains eight ORFs, which encode proteins with collagen motifs. It is not known, if all these eight collagen motifs encoded in the MIMIV genome are expressed within the fibrils on the viral capsid and if these collagenous fibrils are indeed glycosylated. Analysis of the glycans of the MIMIV fibrils would give further information on the type of glycans, if MIMIV collagen glycans also contain Gal as seen in animal collagens and if MIMIV also has collagen disaccharides. This information could also help to explain the bio-

logical relevance of glycosylated fibrils. One possible explanation could be that the MIMIV collagen-like proteins on the capsid might protect the virus, as it could work as a stable protective shield against environmental adverse conditions. The reported facts that MIMIV is resistant to harsh conditions as UV radiation, mechanical stress and elevated temperature corroborate the protective properties of the fibrils.

A difference between the human ColGalT and the MIMIV ColGlcT is its subcellular localization. While GLT25D1 and GLT25D2 are type II transmembrane proteins, L230 is a soluble enzyme. This feature would enable the use of the soluble recombinant L230 enzyme as an improved tool for *in vitro* collagen glycosylation, as a soluble protein is much easier to isolate. Correctly glycosylated collagen is necessary for the medical application of collagen in humans. In fact, the application of bovine collagen and of unmodified recombinant human collagen has been shown to induce allergic reaction in a non-negligible number of treated persons. For this purpose, the soluble L230, which catalyzes the addition of core-Glc, has to be engineered to a ColGalT for use in biotechnology. From blood-group glycosyltransferases it is known that only a few amino acids can change the specificity of a transferase (11, 26). Studying the domain which is responsible for the recognition of UDP-Glc and to change it into a UDP-Gal recognition domain would give rise to a soluble ColGalT which can be used for *in vitro* glycosylation of human collagen.

Acknowledgement

We thank Prof. Dr. D. Raoult (Université de la Méditerranée, CNRS, Marseille, France) for the MIMIV and the *A. polyphaga* cells and Prof. Dr. R. Tenni (University of Pavia, Italy) for the GHyl and GGHyl amino acid standards. This work was supported by Telethon Action Suisse.

Appendix

Supplementary Table A1: HPLC gradient

Time [min]	Solvent A [%]	Solvent B [%]	Solvent C [%]
0.0	75	0	25
74.2	64	0	36
78.7	64	0	36
78.9	0	65	35
104.0	0	65	35
129.3	0	25	75
135.3	0	25	75
153.1	75	0	25
188.9	75	0	25

References

1. **Bank, R. A., B. Beekman, R. Tenni, and J. M. TeKoppele.** 1997. Pre-column derivatisation method for the measurement of glycosylated hydroxylysines of collagenous proteins. J Chromatogr B Biomed Sci Appl **703:**267-272.

2. **Celerin, M., J. M. Ray, N. J. Schisler, A. W. Day, W. G. Stetler-Stevenson, and D. E. Laudenbach.** 1996. Fungal fimbriae are composed of collagen. Embo J **15:**4445-4453.

3. **Cudic, M., D. A. Patel, J. L. Lauer-Fields, K. Brew, and G. B. Fields.** 2008. Development of a convenient peptide-based assay for lysyl hydroxylase. Biopolymers **90:**330-338.

4. **Grassmann, W., and H. Schleich.** 1935. Collagen. II. Carbohydrate content of collagen. Biochem Z **277:**320-328.

5. **Graves, M. V., C. T. Bernadt, R. Cerny, and J. L. Van Etten.** 2001. Molecular and genetic evidence for a virus-encoded glycosyltransferase involved in protein glycosylation. Virology **285:**332-345.

6. **Hennet, T., A. Dinter, P. Kuhnert, T. S. Mattu, P. M. Rudd, and E. G. Berger.** 1998. Genomic cloning and expression of three murine UDP-galactose: beta-N-acetylglucosamine beta1,3-galactosyltransferase genes. J Biol Chem **273:**58-65.

7. **Katzman, R. L., M. H. Halford, V. N. Reinhold, and R. W. Jeanloz.** 1972. Isolation and structure determination of glucosylgalactosylhydroxylysine from sponge and sea anemone collagen. Biochemistry **11:**1161-1167.

8. **Kivirikko, K. I., and R. Myllyla.** 1982. Posttranslational enzymes in the biosynthesis of collagen: intracellular enzymes. Methods Enzymol **82:**245-304.

9. **La Scola, B., S. Audic, C. Robert, L. Jungang, X. de Lamballerie, M. Drancourt, R. Birtles, J. M. Claverie, and D. Raoult.** 2003. A giant virus in amoebae. Science **299:**2033.

10. **Markine-Goriaynoff, N., L. Gillet, J. L. Van Etten, H. Korres, N. Verma, and A. Vanderplasschen.** 2004. Glycosyltransferases encoded by viruses. J Gen Virol **85:**2741-2754.

11. **Patenaude, S. I., N. O. Seto, S. N. Borisova, A. Szpacenko, S. L. Marcus, M. M. Palcic, and S. V. Evans.** 2002. The structural basis for specificity in human ABO(H) blood group biosynthesis. Nat Struct Biol **9:**685-690.

12. **Raoult, D., S. Audic, C. Robert, C. Abergel, P. Renesto, H. Ogata, B. La Scola, M. Suzan, and J. M. Claverie.** 2004. The 1.2-megabase genome sequence of Mimivirus. Science **306:**1344-1350.

13. **Rasmussen, M., and L. Bjorck.** 2001. Unique regulation of SclB - a novel collagen-like surface protein of Streptococcus pyogenes. Mol Microbiol **40:**1427-38.

14. **Rasmussen, M., A. Eden, and L. Bjorck.** 2000. SclA, a novel collagen-like surface protein of *Streptococcus pyogenes*. Infect Immun **68:**6370-6377.

15. **Schegg, B., C. Rutschmann, A. J. Huelsmeier, and T. Hennet.** 2008. Core glycosylation of collagen is initiated by two β(1-O)galactosyltransferases. submitted to Mol. Cell. Biol.

16. **Spiro, M. J., and R. G. Spiro.** 1971. Studies on the biosynthesis of the hydroxylsine-linked disaccharide unit of basement membranes and collagens. II. Kidney galactosyltransferase. J Biol Chem **246:**4910-4918.

17. **Spiro, R. G.** 1969. Characterization and quantitative determination of the hydroxylysine-linked carbohydrate units of several collagens. J Biol Chem **244:**602-612.

18. **Sugrue, R. J.** 2007. Viruses and glycosylation: an overview. Methods Mol Biol **379:**1-13.

19. **Suzan-Monti, M., B. La Scola, and D. Raoult.** 2006. Genomic and evolutionary aspects of Mimivirus. Virus Res **117:**145-155.

20. **Takaluoma, K., J. Lantto, and J. Myllyharju.** 2007. Lysyl hydroxylase 2 is a specific telopeptide hydroxylase, while all three isoenzymes hydroxylate collagenous sequences. Matrix Biol **26:**396-403.

21. **Tenni, R., D. Rimoldi, G. Zanaboni, G. Cetta, and A. A. Castellani.** 1984. Hydroxylysine glycosides: preparation and analysis by reverse phase high performance liquid chromatography. Ital J Biochem **33:**117-127.

22. **Valtavaara, M., C. Szpirer, J. Szpirer, and R. Myllyla.** 1998. Primary structure, tissue distribution, and chromosomal localization of a novel isoform of lysyl hydroxylase (lysyl hydroxylase 3). J Biol Chem **273:**12881-12886.

23. **Xiao, C., P. R. Chipman, A. J. Battisti, V. D. Bowman, P. Renesto, D. Raoult, and M. G. Rossmann.** 2005. Cryo-electron microscopy of the giant Mimivirus. J Mol Biol **353:**493-496.

24. **Xu, Y., D. R. Keene, J. M. Bujnicki, M. Hook, and S. Lukomski.** 2002. Streptococcal Scl1 and Scl2 proteins form collagen-like triple helices. J Biol Chem **277:**27312-27318.

25. **Yamada, T., N. Chuchird, T. Kawasaki, K. Nishida, and S. Hiramatsu.** 1999. *Chlorella* viruses as a source of novel enzymes. J Biosci Bioeng **88:**353-361.

26. **Yazer, M. H.** 2005. What a difference 2 nucleotides make: a short review of ABO genetics. Transfus Med Rev **19:**200-209.

Figure Legends

Fig. 1: Cladogram of human ColGalT related proteins. Human ColGalT GLT25D1 and GLT25D2 as well as the non-active homologue CEECAM1 were blasted against proteins of other species and aligned using the ClustalW algorithm. Homologous proteins were found not only in eukaryotes but also in several bacteria and even in MIMIV.

Fig. 2: Protein Alignment and L230 topology. (A) The protein alignment shows the homology of the MIMIV L230 protein to the human GLT25D1 and GLT25D2 in the N-terminal region and the homology of L230 to the human LH1, LH2 and LH3 in the C-terminal region. (B) L230 has two domains. The N-terminal region, shaded in grey, is homologous to the human ColGalT GLT25D1 and GLT25D2. The C-terminal region shows a homology with the human lysyl hydroxylases LH1, LH2 and LH3.

Fig. 3: Enzymatic activity of MIMIV L230 protein. Mock and L230 transfected Sf9 cells were tested for ColGlcT and ColGalT activity on the acceptor collagen type I. L230 shows a ColGlcT but not a ColGalT activity. Bars indicate the means of six assays. Error bars indicate the SD.

Fig. 4: ColGlcT activity of L230 towards synthetic acceptor peptides. L230 was assayed on the MIMIV collagen peptide L71 (GDVGD-Hyl-GDVGD-

Hyl-GDVGD-Hyl-G), the human Col2A1 (GPTGVTGP-Hyl-KGARGAQGP) and the human ADIPOQ (GDPGLIGP-Hyl-GDIGETGVP). Bars indicate the means of four assays. Error bars indicate the SD.

Fig. 5: ColGlcT reaction product identification. The first panel represents a 17 amino acid standard containing the standards GHyl and GGHyl separated on RP-HPLC. Panel two shows a collagen type II hydrolyzate with the glycosylated hydroxylysine GHyl and GGHyl. The third panel shows the radioactive trace obtained after the reaction of L230 (black) with collagen type I compared with the radioactive trace of the reaction with mock transfected cells (grey). [^{3}H]Val and [^{14}C]Tyr were used as internal amino acid standards. Amino acids are marked in single-letter code.

Fig. 6: ColGlcT enzymatic activity assay with truncated L230. Constructs bearing only the GLT25D-like domain (L230 ΔC-term) or the LH-like domain (L230 ΔN-term) were analyzed for their ColGlcT activity on collagen type I.

Fig. 7: Lysyl hydroxylase activity of L230. The MIMIV L230 and the human LH1 enzymes were assayed on the acceptor peptides $(GIK)_4$ and $(GDK)_4$. Bars indicate the means of four to six assays. Error bars indicate the SD.

Fig. 8: Northern blotting of mimiviral enzymes and collagen. The onset of RNA expression of mimiviral collagens and of enzymes potentially involved in collagen modification was explored during mimiviral infection of *A. polyphaga*. Different infection time points were analyzed: mock, 4, 8, 16, 24 h post infection (hpi). The first panel shows the putative glycosyltransferase R699, the second one L230 and the third panel the viral collagen R241.

Figure 1

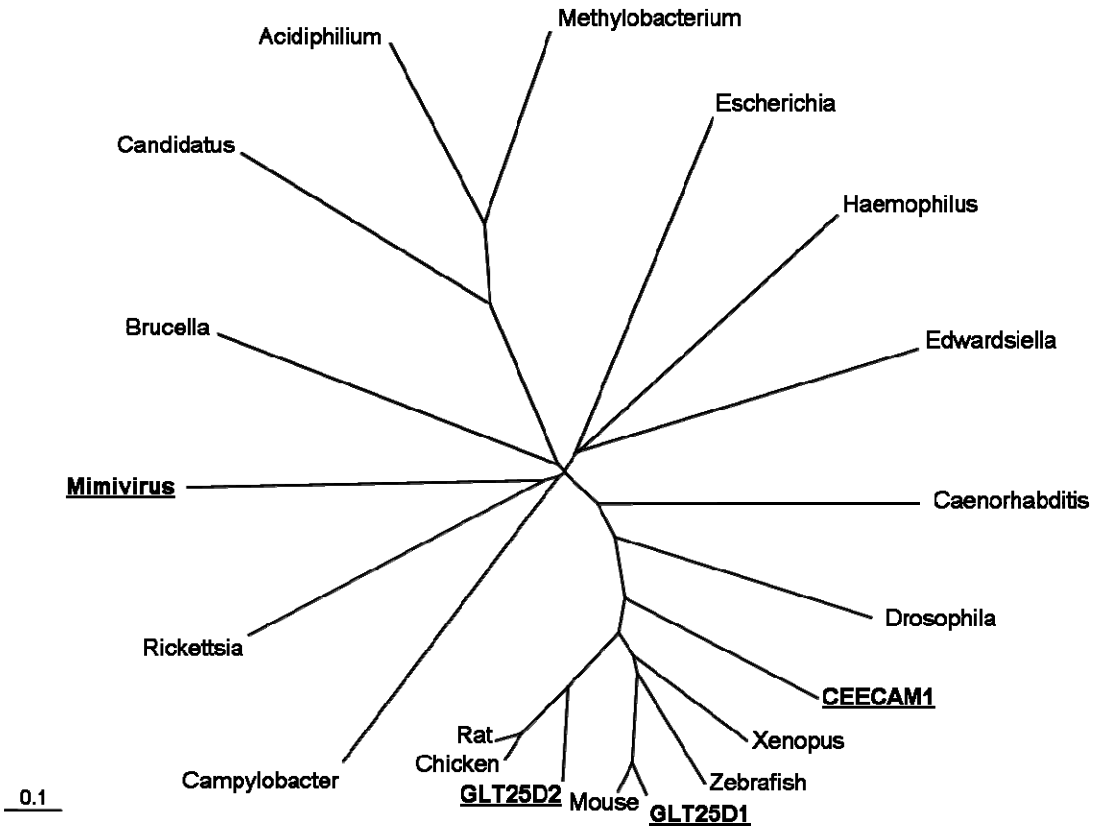
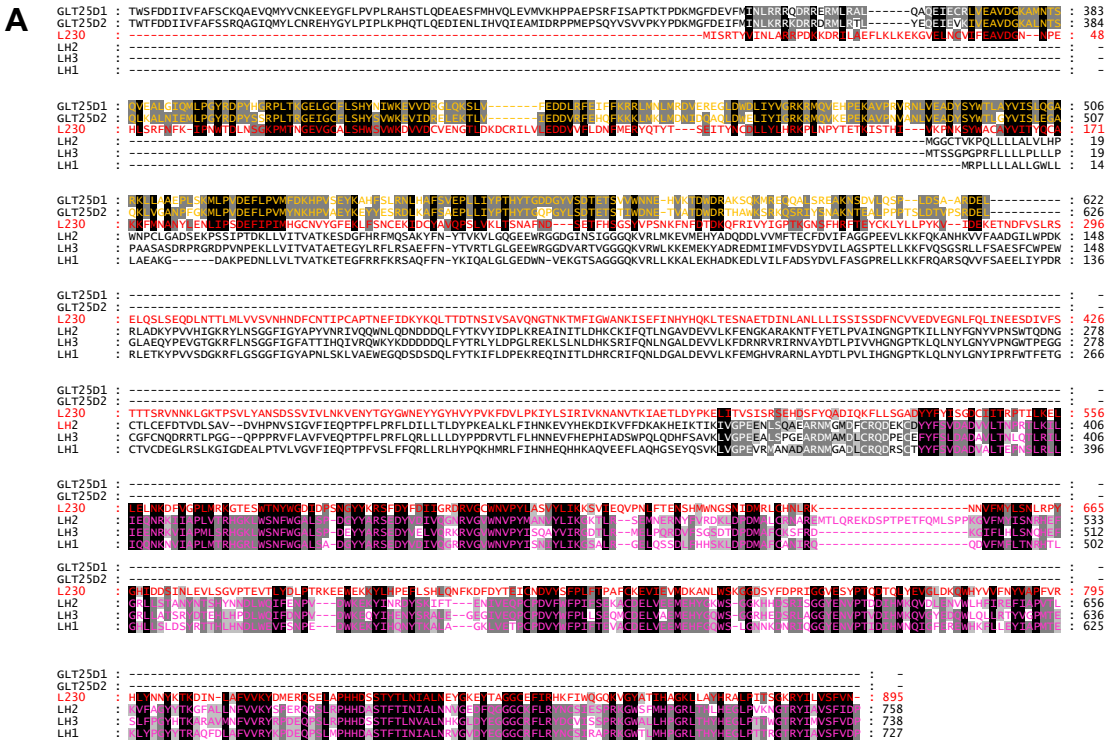

Methylobacterium
Acidiphilium
Escherichia
Candidatus
Haemophilus
Brucella
Edwardsiella
Mimivirus
Caenorhabditis
Drosophila
Rickettsia
CEECAM1
Rat
Xenopus
Chicken
Zebrafish
Campylobacter
GLT25D2
Mouse
GLT25D1
0.1

Figure 2


A
GLT25D1
GLT25D2
L230
LH2
LH3
LH1

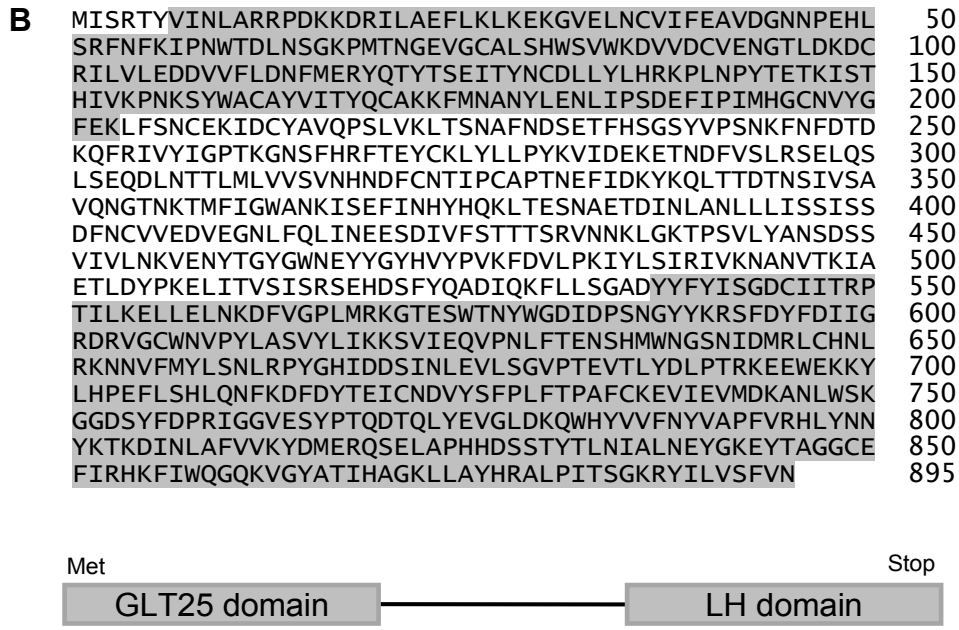

B
MISRTYVINLARRPDKKDRILAEFLKLKEKGVELNCVIFEAVDGNNPEHL 50
SRFNFKIPNWTDLNSGKPMTNGEVGCALSHWSVWKDVVDCVENGTLDKDC 100
RILVLEDDVVFLDNFMERYQTYTSEITYNCDLLYLHRKPLNPYTETKIST 150
HIVKPNKSYWACAYVITYQCAKKFMNANYLENLIPSDEFIPIMHGCNVYG 200
FEKLFSNCEKIDCYAVQPSLVKLTSNAFNDSETFHSGSYVPSNKFNFDTD 250
KQFRIVYIGPTKGNSFHRFTEYCKLYLLPYKVIDEKETNDFVSLRSELQS 300
LSEQDLNTTLMLVVSVNHNDFCNTIPCAPTNEFIDKYKQLTTDTNSIVSA 350
VQNGTNKTMFIGWANKISEFINHYHQKLTESNAETDINLANLLLISSISS 400
DFNCVVEDVEGNLFQLINEESDIVFSTTTSRVNNKLGKTPSVLYANSDSS 450
VIVLNKVENYTGYGWNEYYGYHVYPVKFDVLPKIYLSIRIVKNANVTKIA 500
ETLDYPKELITVSISRSEHDSFYQADIQKFLLSGADYYFYISGDCIITRP 550
TILKELLELNKDFVGPLMRKGTESWTNYWGDIDPSNGYYKRSFDYFDIIG 600
RDRVGCWNVPYLASVYLIKKSVIEQVPNLFTENSHMWNGSNIDMRLCHNL 650
RKNNVFMYLSNLRPYGHIDDSINLEVLSGVPTEVTLYDLPTRKEEWEKKY 700
LHPEFLSHLQNFKDFDYTEICNDVYSFPLFTPAFCKEVIEVMDKANLWSK 750
GGDSYFDPRIGGVESYPTQDTQLYEVGLDKQWHYVVFNYVAPFVRHLYNN 800
YKTKDINLAFVVKYDMERQSELAPHHDSSTYTLNIALNEYGKEYTAGGCE 850
FIRHKFIWQGQKVGYATIHAGKLLAYHRALPITSGKRYILVSFVN 895
Met
Stop
GLT25 domain
LH domain

Figure 3

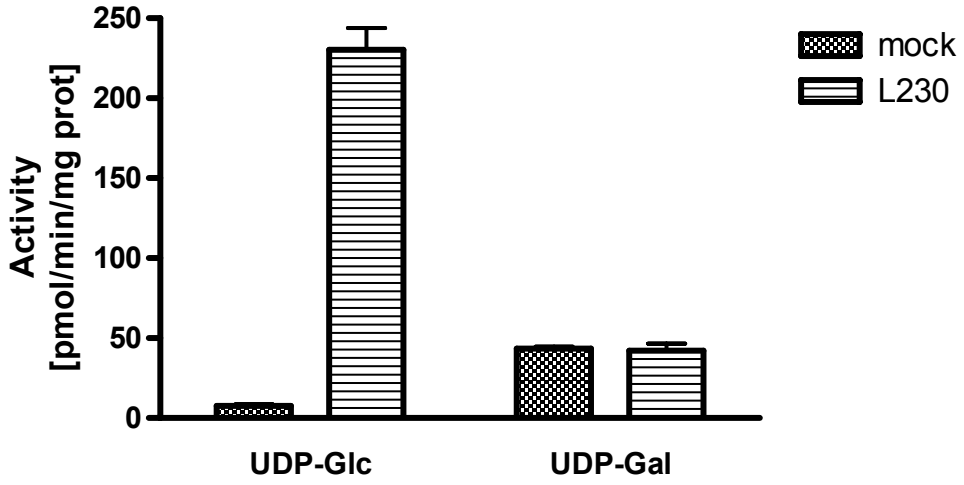

250
200
150
100
50
0
Activity
[pmol/min/mg prot]
mock
L230
UDP-Glc
UDP-Gal

Figure 4

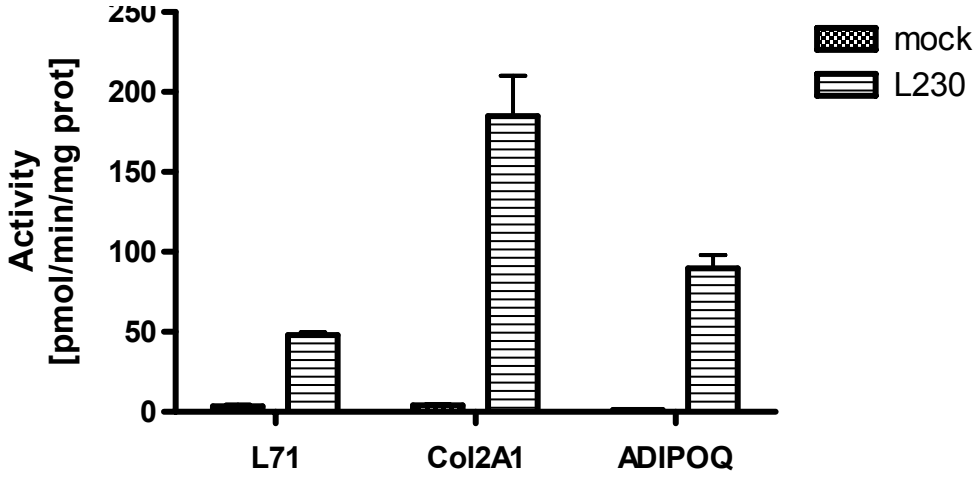
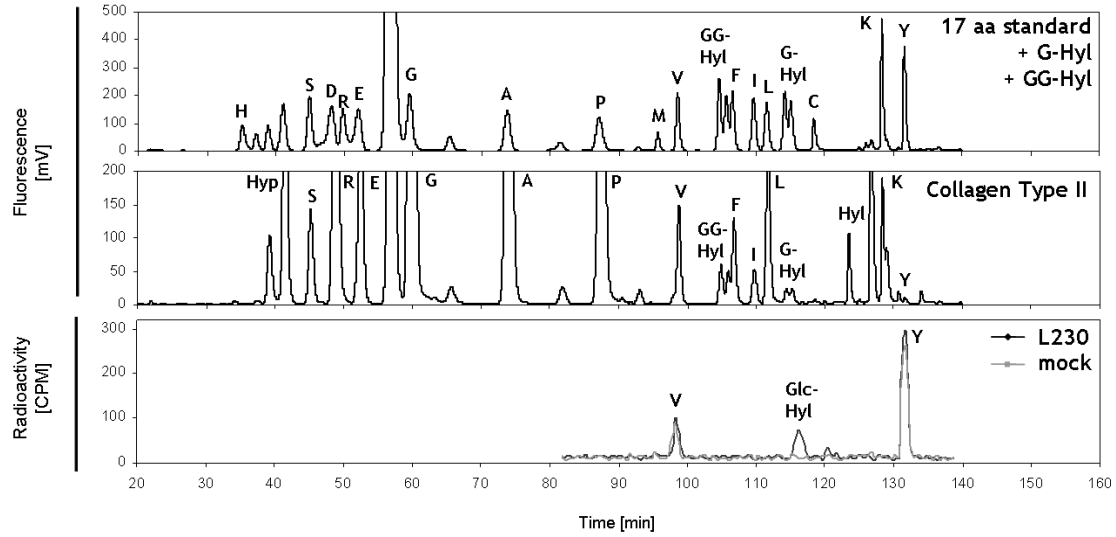

250
200
150
100
50
0
Activity
[pmol/min/mg prot]
L71
Col2A1
ADIPOQ
mock
L230

Figure 5



Hyp: hydroxyproline
Hyl: hydroxylysine

G-Hyl: Gal-hydroxylysine
GG-Hyl: Glc-Gal-hydroxylysine

Figure 6

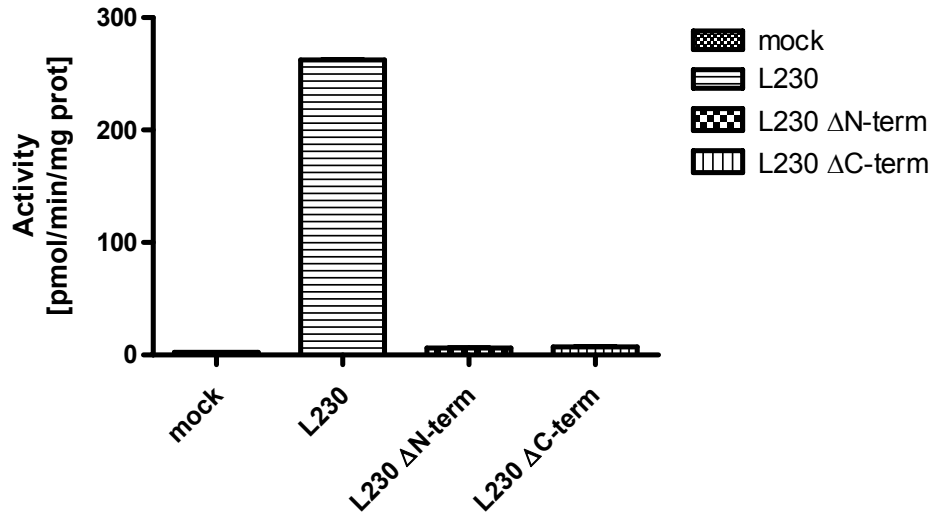

Activity
[pmol/min/mg prot]
300
200
100
0
mock
L230
L230 ΔN-term
L230 ΔC-term
mock
L230
L230 ΔN-term
L230 ΔC-term

Figure 7

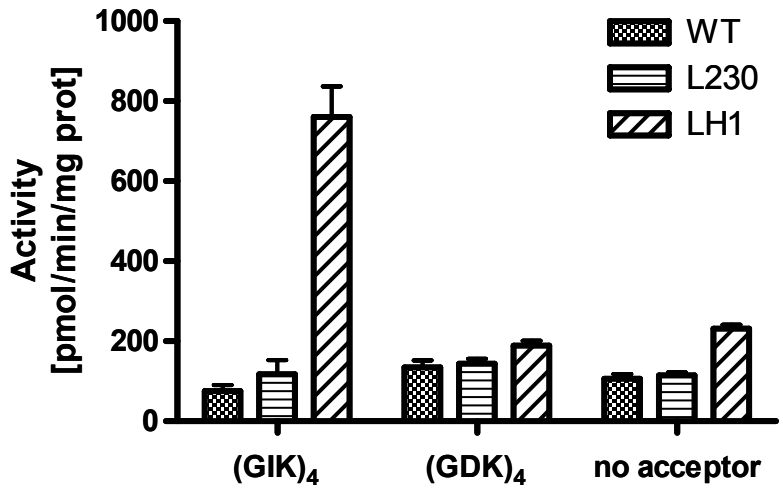
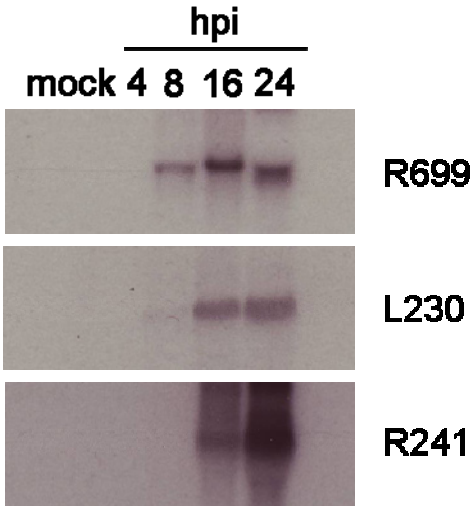

1000
800
600
400
200
0
Activity
[pmol/min/mg prot]
WT
L230
LH1
(GIK)4
(GDK)4
no acceptor

Figure 8



3 Discussion

Glycosylation is a ubiquitous modification of collagen. Up to now, only the LH3 enzyme has been reported to mediate collagen glycosylation quite inefficiently. In the present work, we have identified two collagen glycosyltransferase enzymes, GLT25D1 and GLT25D2, as well as the non-active protein CEECAM1, which share a high degree of sequence identity to each other.

Tissue Northern Blotting of the human ColGalT GLT25D2 showed that the expression of this enzyme is confined to neuronal tissue. As neuronal tissue only expresses collagen type IV in small amounts in the cerebrovascular basal lamina as part of the blood brain barrier (126-128), and collagen XVII as possible anchor of neurons to their substratum in various neuronal regions (129) the question arises on the main substrate of GLT25D2 in the brain. As GLT25D2 is also able to glycosylate non-collagenous proteins, which contain a collagen like domain, acetylcholinesterase could be a possible substrate. Acetylcholinesterase is responsible for the rapid hydrolysis of the neurotransmitter acetylcholine at nicotinic cholinergic synapses terminating muscle contraction. The asymmetric acetylcholinesterase consists of one to three catalytic tetramers linked to a collagen-like tail, the ColQ protein. ColQ contains a collagen-like domain, which is characterized by Gly-X-Y repeats and forms a collagen triple helix. The collagen-tail of acetylcholinesterase is responsible for the anchoring of the enzyme to the basal lamina of the neuromuscular junction (130-132). The ColQ of acetylcholinesterase, which is located at the neuronal junction, could be a good target for the GLT25D2 activity. This would mean that beside collagens expressed in the neuronal tissue, also non-collagenous proteins could be targets of GLT25D2.

As collagen functions as a central unit of the connective tissue, the question of the presence of collagen in simple organisms is of interest. Not a long time ago, MIMIV has been detected as a unique organism being clearly a virus but also showing features never seen before in viruses. In the present work, we have characterized the mimiviral ColGlcT L230 protein. It appears that collagen glycosylation in MIMIV is different than in animal collagens. In animal, collagens either Gal or glucosylgalactose are attached to Hyl but in our *in*

vitro enzymatic activity assay, we could glucosylate Hyl in different collagen molecules, which was never described up to now. It was suggested that MIMIV virions also consist of collagenous proteins (133, 134). The fibrils surrounding the virus capsid might be post-translationally modified by virus encoded enzymes. These glycosylated capsid proteins possibly contribute to the high resistance of MIMIV to harsh conditions outside the host cell, like elevated temperature, mechanical stress and UV radiation.

Even though the MIMIV ColGlcT has been identified based on sequence homology to the human ColGalT, the mimiviral ColGlcT catalyzes the transfer of UDP-Glc and not UDP-Gal onto collagen as would be expected. From the ABO blood group glycosyltransferases, it is known that only a few amino acids can determine the substrate specificity of a glycosyltransferase (135). The GalNAc-transferase of the blood group A and the Gal-transferase of the blood group B differ in only four amino acids on the whole sequence. Of these four amino acids only one single amino acid residue is responsible for the biosynthesis of the different carbohydrates of blood group A and blood group B (135). As only one single amino acid can determine the substrate specificity of a glycosyltransferase, this could explain the different properties of the MIMIV ColGlcT L230 and of the human ColGalT, although they share a high degree of sequence similarity.

Currently, the function of the collagen glycosylation is not understood, although the glycan structure of collagen was already solved in 1967 (95). The glycans might be involved in the lateral packing of the collagen molecule into fibrils, which might influence the diameter of the collagen fibrils (102). The glycans may play an important role in the secretion of the collagen molecules, the glycans may act as a targeting signal for proper secretion of the collagen fibrils into the extracellular matrix. This was demonstrated for the non-collagenous proteins ADIPOQ and MBL, which are not properly oligomerized and secreted (25, 103). However, to completely understand the function of collagen glycosylation, it is important to identify all enzymes involved in this process. The identification and characterization of the two human ColGalTs GLT25D1 and GLT25D2 is the first step in understanding collagen glycosylation and it might facilitate genetic investigations to further elaborate the role of collagen glycosylation, also in respect to connective tis-

sue disorders. But to fully understand collagen glycosylation, the second enzyme, the ColGlcT, which is also involved in this process, has to be identified as well. Based on homology blast searches against known α-glucosyltransferases several candidates for human ColGlcTs were already tested (data not shown). However, as the ColGlcT candidates did not show any ColGlcT activity, the human ColGlcT still remains unknown.

4 Abbreviations

3Hyp	3-hydroxyproline
4Hyp	4-hydroxyproline
ADIPOQ	adiponectin
BCS	Brittle Cornea Syndrome
CDG	congenital disorders of glycosylation
CMP	cytidine monophosphate
ColGalT	Collagen galactosyltransferase
ColGlcT	Collagen glucosyltransferase
CRD	carbohydrate recognition domain
EDS	Ehlers-Danlos-Syndrome
EGF	epidermal growth factor
ER	Endoplasmic Reticulum
Fuc	fucose
Gal	galactose
GalNAc	*N*-acetylgalactosamin
GalT	galactosyltransferase
Glc	glucose
GlcA	glucuronic acid
GlcNAc	*N*-acetylglucosamin
GMP	guanosine monophosphate
GT	glycosyltransferase
HMW	high molecular weight
Hyl	Hydroxylysine
Hyp	Hydroxyproline
LH	lysylhydroxylase
LLO	lipid-linked oligosaccharide
Man	mannose
MBL	Mannose binding lectin
NDP	nucleotide diphosphate
NMP	nucleotide monophosphate
OI	Osteogenesis Imperfecta

P3H	prolyl-3-hydroxylase
P4H	prolyl-4-hydroxylase
Sia	sialic acid
UDP	uridine diphosphate
UDP	uridine monophosphate
Xyl	Xylose

5 References

1. Vliegenthart, J. F., and Casset, F. (1998) Novel forms of protein glycosylation. *Curr Opin Struct Biol* 8, 565-571
2. Furmanek, A., and Hofsteenge, J. (2000) Protein C-mannosylation: facts and questions. *Acta Biochim Pol* 47, 781-789
3. Helenius, A., and Aebi, M. (2001) Intracellular functions of N-linked glycans. *Science* 291, 2364-2369
4. Helenius, A., and Aebi, M. (2004) Roles of N-linked glycans in the endoplasmic reticulum. *Annu Rev Biochem* 73, 1019-1049
5. Wopereis, S., Lefeber, D. J., Morava, E., and Wevers, R. A. (2006) Mechanisms in protein O-glycan biosynthesis and clinical and molecular aspects of protein O-glycan biosynthesis defects: a review. *Clin Chem* 52, 574-600
6. Peter-Katalinic, J. (2005) Methods in enzymology: O-glycosylation of proteins. *Methods Enzymol* 405, 139-171
7. Chen, H., Wang, Z., Sun, Z., Kim, E. J., and Yarema, K. J. (2005) *Mammalian glycosylation: An overview of carbohydrate glycosylation*, CRC Press, Taylor & Francis Group
8. Van den Steen, P., Rudd, P. M., Dwek, R. A., and Opdenakker, G. (1998) Concepts and principles of O-linked glycosylation. *Crit Rev Biochem Mol Biol* 33, 151-208
9. Luo, Y., and Haltiwanger, R. S. (2005) O-fucosylation of notch occurs in the endoplasmic reticulum. *J Biol Chem* 280, 11289-11294
10. Krusius, T., Finne, J., Margolis, R. K., and Margolis, R. U. (1986) Identification of an O-glycosidic mannose-linked sialylated tetrasaccharide and keratan sulfate oligosaccharides in the chondroitin sulfate proteoglycan of brain. *J Biol Chem* 261, 8237-8242
11. O'Donnell, N. (2002) Intracellular glycosylation and development. *Biochim Biophys Acta* 1573, 336-345
12. Hirschberg, C. B., Robbins, P. W., and Abeijon, C. (1998) Transporters of nucleotide sugars, ATP, and nucleotide sulfate in the endoplasmic reticulum and Golgi apparatus. *Annu Rev Biochem* 67, 49-69
13. Handford, M., Rodriguez-Furlan, C., and Orellana, A. (2006) Nucleotide-sugar transporters: structure, function and roles in vivo. *Braz J Med Biol Res* 39, 1149-1158
14. Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Hart, G. W., and Marth, J. D. (1999) *Essentials of Glycobiology*, CSHL Press: La Jolla, California USA
15. Aebi, M., and Hennet, T. (2001) Congenital disorders of glycosylation: genetic model systems lead the way. *Trends Cell Biol* 11, 136-141
16. Freeze, H. H. (2006) Genetic defects in the human glycome. *Nat Rev Genet* 7, 537-551
17. Myllyharju, J., and Kivirikko, K. I. (2004) Collagens, modifying enzymes and their mutations in humans, flies and worms. *Trends Genet* 20, 33-43

18. Soderhall, C., Marenholz, I., Kerscher, T., Ruschendorf, F., Esparza-Gordillo, J., Worm, M., Gruber, C., Mayr, G., Albrecht, M., Rohde, K., Schulz, H., Wahn, U., Hubner, N., and Lee, Y. A. (2007) Variants in a novel epidermal collagen gene (COL29A1) are associated with atopic dermatitis. *PLoS Biol* 5, 19521961
19. Veit, G., Kobbe, B., Keene, D. R., Paulsson, M., Koch, M., and Wagener, R. (2006) Collagen XXVIII, a novel von Willebrand factor A domain-containing protein with many imperfections in the collagenous domain. *J Biol Chem* 281, 3494-3504
20. Gelse, K., Poschl, E., and Aigner, T. (2003) Collagens: structure, function, and biosynthesis. *Adv Drug Deliv Rev* 55, 1531-1546
21. Royce, P. M., and Steinmann, B. (2002) *Connective Tissue and its heritable disorders. Molecular, genetic and medical aspects*, Wiley-Liss
22. Digtyar, A. V., Pozdnyakova, N. V., Feldman, N. B., Lutsenko, S. V., and Severin, S. E. (2007) Endostatin: current concepts about its biological role and mechanisms of action. *Biochemistry (Mosc)* 72, 235-246
23. Eyre, D. R., Paz, M. A., and Gallop, P. M. (1984) Cross-linking in collagen and elastin. *Annu Rev Biochem* 53, 717-748
24. Franzke, C. W., Bruckner, P., and Bruckner-Tuderman, L. (2005) Collagenous transmembrane proteins: recent insights into biology and pathology. *J Biol Chem* 280, 4005-4008
25. Richards, A. A., Stephens, T., Charlton, H. K., Jones, A., Macdonald, G. A., Prins, J. B., and Whitehead, J. P. (2006) Adiponectin multimerization is dependent on conserved lysines in the collagenous domain: evidence for regulation of multimerization by alterations in posttranslational modifications. *Mol Endocrinol* 20, 1673-1687
26. Hakansson, K., and Reid, K. B. (2000) Collectin structure: a review. *Protein Sci* 9, 1607-1617
27. Shinkai, H., and Yonemasu, K. (1979) Hydroxylysine-linked glycosides of human complement subcomponent C1q and various collagens. *Biochem J* 177, 847-852
28. Bon, S., Ayon, A., Leroy, J., and Massoulie, J. (2003) Trimerization domain of the collagen tail of acetylcholinesterase. *Neurochem Res* 28, 523-535
29. Wang, Y., Lam, K. S., Yau, M. H., and Xu, A. (2008) Post-translational modifications of adiponectin: mechanisms and functional implications. *Biochem J* 409, 623-633
30. van de Wetering, J. K., van Golde, L. M., and Batenburg, J. J. (2004) Collectins: players of the innate immune system. *Eur J Biochem* 271, 1229-1249
31. Sastry, K., and Ezekowitz, R. A. (1993) Collectins: pattern recognition molecules involved in first line host defense. *Curr Opin Immunol* 5, 59-66
32. Ogasawara, Y., and Voelker, D. R. (1995) Altered carbohydrate recognition specificity engineered into surfactant protein D reveals different binding mechanisms for phosphatidylinositol and glucosylceramide. *J Biol Chem* 270, 14725-14732

33. Thiel, S. (2007) Complement activating soluble pattern recognition molecules with collagen-like regions, mannan-binding lectin, ficolins and associated proteins. *Mol Immunol* 44, 3875-3888
34. Degn, S. E., Thiel, S., and Jensenius, J. C. (2007) New perspectives on mannan-binding lectin-mediated complement activation. *Immunobiology* 212, 301-311
35. Arnold, J. N., Dwek, R. A., Rudd, P. M., and Sim, R. B. (2006) Mannan binding lectin and its interaction with immunoglobulins in health and in disease. *Immunol Lett* 106, 103-110
36. Kishore, U., Greenhough, T. J., Waters, P., Shrive, A. K., Ghai, R., Kamran, M. F., Bernal, A. L., Reid, K. B., Madan, T., and Chakraborty, T. (2006) Surfactant proteins SP-A and SP-D: structure, function and receptors. *Mol Immunol* 43, 1293-1315
37. Crouch, E., and Wright, J. R. (2001) Surfactant proteins a and d and pulmonary host defense. *Annu Rev Physiol* 63, 521-554
38. Wright, J. R. (2005) Immunoregulatory functions of surfactant proteins. *Nat Rev Immunol* 5, 58-68
39. Mizuno, K., Hayashi, T., and Bachinger, H. P. (2003) Hydroxylation-induced stabilization of the collagen triple helix. Further characterization of peptides with 4(R)-hydroxyproline in the Xaa position. *J Biol Chem* 278, 32373-32379
40. Vranka, J. A., Sakai, L. Y., and Bachinger, H. P. (2004) Prolyl 3-hydroxylase 1, enzyme characterization and identification of a novel family of enzymes. *J Biol Chem* 279, 23615-23621
41. Risteli, M., Niemitalo, O., Lankinen, H., Juffer, A. H., and Myllyla, R. (2004) Characterization of collagenous peptides bound to lysyl hydroxylase isoforms. *J Biol Chem* 279, 37535-37543
42. Valtavaara, M., Szpirer, C., Szpirer, J., and Myllyla, R. (1998) Primary structure, tissue distribution, and chromosomal localization of a novel isoform of lysyl hydroxylase (lysyl hydroxylase 3). *J Biol Chem* 273, 12881-12886
43. Kivirikko, K. I., and Myllyla, R. (1985) Post-translational processing of procollagens. *Ann N Y Acad Sci* 460, 187-201
44. Koivu, J. (1987) Identification of disulfide bonds in carboxy-terminal propeptides of human type I procollagen. *FEBS Lett* 212, 229-232
45. Lees, J. F., and Bulleid, N. J. (1994) The role of cysteine residues in the folding and association of the COOH-terminal propeptide of types I and III procollagen. *J Biol Chem* 269, 24354-24360
46. Zafarullah, K., Brown, E. M., Kuivaniemi, H., Tromp, G., Sieron, A. L., Fertala, A., and Prockop, D. J. (1997) Synthesis and conformational properties of a recombinant C-propeptide of human type III procollagen. *Matrix Biol* 16, 201-209
47. Lees, J. F., Tasab, M., and Bulleid, N. J. (1997) Identification of the molecular recognition sequence which determines the type-specific assembly of procollagen. *Embo J* 16, 908-916
48. Lamande, S. R., and Bateman, J. F. (1999) Procollagen folding and assembly: the role of endoplasmic reticulum enzymes and molecular chaperones. *Semin Cell Dev Biol* 10, 455-464
49. Bonfanti, L., Mironov, A. A., Jr., Martinez-Menarguez, J. A., Martella, O., Fusella, A., Baldassarre, M., Buccione, R., Geuze, H. J., Mironov,

A. A., and Luini, A. (1998) Procollagen traverses the Golgi stack without leaving the lumen of cisternae: evidence for cisternal maturation. *Cell* 95, 993-1003
50. Trelstad, R. L., and Hayashi, K. (1979) Tendon collagen fibrillogenesis: intracellular subassemblies and cell surface changes associated with fibril growth. *Dev Biol* 71, 228-242
51. Tasab, M., Jenkinson, L., and Bulleid, N. J. (2002) Sequence-specific recognition of collagen triple helices by the collagen-specific molecular chaperone HSP47. *J Biol Chem* 277, 35007-35012
52. Koide, T., Asada, S., Takahara, Y., Nishikawa, Y., Nagata, K., and Kitagawa, K. (2006) Specific recognition of the collagen triple helix by chaperone HSP47: minimal structural requirement and spatial molecular orientation. *J Biol Chem* 281, 3432-3438
53. Starkuviene, V., and Pepperkok, R. (2007) Differential requirements for ts-O45-G and procollagen biosynthetic transport. *Traffic* 8, 1035-1051
54. Mironov, A. A., Beznoussenko, G. V., Nicoziani, P., Martella, O., Trucco, A., Kweon, H. S., Di Giandomenico, D., Polishchuk, R. S., Fusella, A., Lupetti, P., Berger, E. G., Geerts, W. J., Koster, A. J., Burger, K. N., and Luini, A. (2001) Small cargo proteins and large aggregates can traverse the Golgi by a common mechanism without leaving the lumen of cisternae. *J Cell Biol* 155, 1225-1238
55. Prockop, D. J., and Kivirikko, K. I. (1995) Collagens: molecular biology, diseases, and potentials for therapy. *Annu Rev Biochem* 64, 403-434
56. Rhoads, R. E., Udenfriend, S., and Bornstein, P. (1971) In vitro enzymatic hydroxylation of prolyl residues in the alpha 1-CB2 fragment of rat collagen. *J Biol Chem* 246, 4138-4142
57. Berg, R. A., and Prockop, D. J. (1973) The thermal transition of a non-hydroxylated form of collagen. Evidence for a role for hydroxyproline in stabilizing the triple-helix of collagen. *Biochem Biophys Res Commun* 52, 115-120
58. Kivirikko, K. I., Ryhanen, L., Anttinen, H., Bornstein, P., and Prockop, D. J. (1973) Further hydroxylation of lysyl residues in collagen by protocollagen lysyl hydroxylase in vitro. *Biochemistry* 12, 4966-4971
59. Kivirikko, K. I., and Myllyharju, J. (1998) Prolyl 4-hydroxylases and their protein disulfide isomerase subunit. *Matrix Biol* 16, 357-368
60. Myllyharju, J. (2003) Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis. *Matrix Biol* 22, 15-24
61. Kivirikko, K. I., Myllyla, R., and Pihlajaniemi, T. (1992) *Hydroxylation of proline and lysine residues in collagens and other animal and plant proteins*, CRC Press
62. Brodsky, B., and Ramshaw, J. A. (1997) The collagen triple-helix structure. *Matrix Biol* 15, 545-554
63. Kivirikko, K. I., and Myllyla, R. (1982) Posttranslational enzymes in the biosynthesis of collagen: intracellular enzymes. *Methods Enzymol* 82, 245-304
64. Kukkola, L., Hieta, R., Kivirikko, K. I., and Myllyharju, J. (2003) Identification and characterization of a third human, rat, and mouse

collagen prolyl 4-hydroxylase isoenzyme. *J Biol Chem* 278, 47685-47693

65. Annunen, P., Helaakoski, T., Myllyharju, J., Veijola, J., Pihlajaniemi, T., and Kivirikko, K. I. (1997) Cloning of the human prolyl 4-hydroxylase alpha subunit isoform alpha(II) and characterization of the type II enzyme tetramer. The alpha(I) and alpha(II) subunits do not form a mixed alpha(I)alpha(II)beta2 tetramer. *J Biol Chem* 272, 17342-17348
66. Helaakoski, T., Vuori, K., Myllyla, R., Kivirikko, K. I., and Pihlajaniemi, T. (1989) Molecular cloning of the alpha-subunit of human prolyl 4-hydroxylase: the complete cDNA-derived amino acid sequence and evidence for alternative splicing of RNA transcripts. *Proc Natl Acad Sci USA* 86, 4392-4396
67. Pihlajaniemi, T., Helaakoski, T., Tasanen, K., Myllyla, R., Huhtala, M. L., Koivu, J., and Kivirikko, K. I. (1987) Molecular cloning of the beta-subunit of human prolyl 4-hydroxylase. This subunit and protein disulphide isomerase are products of the same gene. *Embo J* 6, 643-649
68. Annunen, P., Autio-Harmainen, H., and Kivirikko, K. I. (1998) The novel type II prolyl 4-hydroxylase is the main enzyme form in chondrocytes and capillary endothelial cells, whereas the type I enzyme predominates in most cells. *J Biol Chem* 273, 5989-5992
69. Nissi, R., Autio-Harmainen, H., Marttila, P., Sormunen, R., and Kivirikko, K. I. (2001) Prolyl 4-hydroxylase isoenzymes I and II have different expression patterns in several human tissues. *J Histochem Cytochem* 49, 1143-1153
70. Vuori, K., Pihlajaniemi, T., Marttila, M., and Kivirikko, K. I. (1992) Characterization of the human prolyl 4-hydroxylase tetramer and its multifunctional protein disulfide-isomerase subunit synthesized in a baculovirus expression system. *Proc Natl Acad Sci USA* 89, 7467-7470
71. Vuorela, A., Myllyharju, J., Nissi, R., Pihlajaniemi, T., and Kivirikko, K. I. (1997) Assembly of human prolyl 4-hydroxylase and type III collagen in the yeast pichia pastoris: formation of a stable enzyme tetramer requires coexpression with collagen and assembly of a stable collagen requires coexpression with prolyl 4-hydroxylase. *Embo J* 16, 6702-6712
72. Myllyharju, J. (2005) Intracellular post-translational modifications of collagens. *Top Curr Chem* 247, 115-147
73. Risteli, J., Tryggvason, K., and Kivirikko, K. I. (1977) Prolyl 3-hydroxylase: partial characterization of the enzyme from rat kidney cortex. *Eur J Biochem* 73, 485-492
74. Wassenhove-McCarthy, D. J., and McCarthy, K. J. (1999) Molecular characterization of a novel basement membrane-associated proteoglycan, leprecan. *J Biol Chem* 274, 25004-125007
75. Kaul, S. C., Sugihara, T., Yoshida, A., Nomura, H., and Wadhwa, R. (2000) Gros1, a potential growth suppressor on chromosome 1: its identity to basement membrane-associated proteoglycan, leprecan. *Oncogene* 19, 3576-3583

76. Morello, R., Bertin, T. K., Chen, Y., Hicks, J., Tonachini, L., Monticone, M., Castagnola, P., Rauch, F., Glorieux, F. H., Vranka, J., Bachinger, H. P., Pace, J. M., Schwarze, U., Byers, P. H., Weis, M., Fernandes, R. J., Eyre, D. R., Yao, Z., Boyce, B. F., and Lee, B. (2006) CRTAP is required for prolyl 3- hydroxylation and mutations cause recessive osteogenesis imperfecta. *Cell* 127, 291-304
77. Cabral, W. A., Chang, W., Barnes, A. M., Weis, M., Scott, M. A., Leikin, S., Makareeva, E., Kuznetsova, N. V., Rosenbaum, K. N., Tifft, C. J., Bulas, D. I., Kozma, C., Smith, P. A., Eyre, D. R., and Marini, J. C. (2007) Prolyl 3-hydroxylase 1 deficiency causes a recessive metabolic bone disorder resembling lethal/severe osteogenesis imperfecta. *Nat Genet* 39, 359-365
78. Marini, J. C., Cabral, W. A., Barnes, A. M., and Chang, W. (2007) Components of the collagen prolyl 3-hydroxylation complex are crucial for normal bone development. *Cell Cycle* 6, 1675-1681
79. Takaluoma, K., Lantto, J., and Myllyharju, J. (2007) Lysyl hydroxylase 2 is a specific telopeptide hydroxylase, while all three isoenzymes hydroxylate collagenous sequences. *Matrix Biol* 26, 396-403
80. Valtavaara, M., Papponen, H., Pirttila, A. M., Hiltunen, K., Helander, H., and Myllyla, R. (1997) Cloning and characterization of a novel human lysyl hydroxylase isoform highly expressed in pancreas and muscle. *J Biol Chem* 272, 6831-6834
81. Passoja, K., Rautavuoma, K., Ala-Kokko, L., Kosonen, T., and Kivirikko, K. I. (1998) Cloning and characterization of a third human lysyl hydroxylase isoform. *Proc Natl Acad Sci USA* 95, 10482-10486
82. Myllyla, R., Pihlajaniemi, T., Pajunen, L., Turpeenniemi-Hujanen, T., and Kivirikko, K. I. (1991) Molecular cloning of chick lysyl hydroxylase. Little homology in primary structure to the two types of subunit of prolyl 4-hydroxylase. *J Biol Chem* 266, 2805-2810
83. Ruotsalainen, H., Sipila, L., Kerkela, E., Pospiech, H., and Myllyla, R. (1999) Characterization of cDNAs for mouse lysyl hydroxylase 1, 2 and 3, their phylogenetic analysis and tissue-specific expression in the mouse. *Matrix Biol* 18, 325-329
84. Heikkinen, J., Hautala, T., Kivirikko, K. I., and Myllyla, R. (1994) Structure and expression of the human lysyl hydroxylase gene (PLOD): introns 9 and 16 contain Alu sequences at the sites of recombination in Ehlers-Danlos syndrome type VI patients. *Genomics* 24, 464-471
85. Myllyla, R., Pajunen, L., and Kivirikko, K. I. (1988) Polyclonal and monoclonal antibodies to human lysyl hydroxylase and studies on the molecular heterogeneity of the enzyme. *Biochem J* 253, 489-496
86. van der Slot, A. J., Zuurmond, A. M., Bardoel, A. F., Wijmenga, C., Pruijs, H. E., Sillence, D. O., Brinckmann, J., Abraham, D. J., Black, C. M., Verzijl, N., DeGroot, J., Hanemaaijer, R., TeKoppele, J. M., Huizinga, T. W., and Bank, R. A. (2003) Identification of PLOD2 as telopeptide lysyl hydroxylase, an important enzyme in fibrosis. *J Biol Chem* 278, 40967-40972
87. Yeowell, H. N., and Walker, L. C. (1999) Tissue specificity of a new splice form of the human lysyl hydroxylase 2 gene. *Matrix Biol* 18, 179-187

88. Myllyla, R., Wang, C., Heikkinen, J., Juffer, A., Lampela, O., Risteli, M., Ruotsalainen, H., Salo, A., and Sipila, L. (2007) Expanding the lysyl hydroxylase toolbox: new insights into the localization and activities of lysyl hydroxylase 3 (LH3). *J Cell Physiol* 212, 323-329
89. Rautavuoma, K., Takaluoma, K., Sormunen, R., Myllyharju, J., Kivirikko, K. I., and Soininen, R. (2004) Premature aggregation of type IV collagen and early lethality in lysyl hydroxylase 3 null mice. *Proc Natl Acad Sci USA* 101, 14120-14125
90. Wang, C., Luosujarvi, H., Heikkinen, J., Risteli, M., Uitto, L., and Myllyla, R. (2002) The third activity for lysyl hydroxylase 3: galactosylation of hydroxylysyl residues in collagens in vitro. *Matrix Biol* 21, 559-566
91. Heikkinen, J., Risteli, M., Wang, C., Latvala, J., Rossi, M., Valtavaara, M., and Myllyla, R. (2000) Lysyl hydroxylase 3 is a multifunctional protein possessing collagen glucosyltransferase activity. *J Biol Chem* 275, 36158-36163
92. Suokas, M., Lampela, O., Juffer, A. H., Myllyla, R., and Kellokumpu, S. (2003) Retrieval-independent localization of lysyl hydroxylase in the endoplasmic reticulum via a peptide fold in its iron-binding domain. *Biochem J* 370, 913-920
93. Kivirikko, K. I., and Pihlajaniemi, T. (1998) Collagen hydroxylases and the protein disulfide isomerase subunit of prolyl 4-hydroxylases. *Adv Enzymol Relat Areas Mol Biol* 72, 325-398
94. Grassmann, W., and Schleich, H. (1935) Collagen. II. Carbohydrate content of collagen. *Biochem Z* 277, 320-328
95. Spiro, R. G. (1967) The structure of the disaccharide unit of the renal glomerular basement membrane. *J Biol Chem* 242, 4813-4823
96. Spiro, M. J., and Spiro, R. G. (1971) Studies on the biosynthesis of the hydroxylsine-linked disaccharide unit of basement membranes and collagens. II. Kidney galactosyltransferase. *J Biol Chem* 246, 4910-4918
97. Kivirikko, K. I., and Myllyla, R. (1979) Collagen glycosyltransferases. *Int Rev Connect Tissue Res* 8, 23-72
98. Brownell, A. G., and Veis, A. (1975) The intracellular location of the glycosylation of hydroxylysine of collagen. *Biochem Biophys Res Commun* 63, 371-377
99. Harwood, R., Grant, M. E., and Jackson, D. S. (1975) Studies on the glycosylation of hydroxylysine residues during collagen biosynthesis and the subcellular localization of collagen galactosyltransferase and collagen glucosyltransferase in tendon and cartilage cells. *Biochem J* 152, 291-302
100. Risteli, L., Myllya, R., and Kivirikko, K. I. (1976) Partial purification and characterization of collagen galactosyltransferase from chick embryos. *Biochem J* 155, 145-153
101. Myllyla, R., Risteli, L., and Kivirikko, K. I. (1975) Glucosylation of galactosylhydroxylysyl residues in collagen in vitro by collagen glucosyltransferase. Inhibition by triple-helical conformation of the substrate. *Eur J Biochem* 58, 517-521
102. Notbohm, H., Nokelainen, M., Myllyharju, J., Fietzek, P. P., Muller, P. K., and Kivirikko, K. I. (1999) Recombinant human type II collagens

with low and high levels of hydroxylysine and its glycosylated forms show marked differences in fibrillogenesis in vitro. *J Biol Chem* 274, 8988-8992
103. Colley, K. J., and Baenziger, J. U. (1987) Post-translational modifications of the core-specific lectin. Relationship to assembly, ligand binding, and secretion. *J Biol Chem* 262, 10296-10303
104. Rautavuoma, K., Takaluoma, K., Passoja, K., Pirskanen, A., Kvist, A. P., Kivirikko, K. I., and Myllyharju, J. (2002) Characterization of three fragments that constitute the monomers of the human lysyl hydroxylase isoenzymes 1-3. The 30-kDa N-terminal fragment is not required for lysyl hydroxylase activity. *J Biol Chem* 277, 23084-23091
105. Myllyla, R., Anttinen, H., Risteli, L., and Kivirikko, K. I. (1977) Isolation of collagen glucosyltransferase as a homogeneous protein from chick embryos. *Biochim Biophys Acta* 480, 113-121
106. Wang, C., Risteli, M., Heikkinen, J., Hussa, A. K., Uitto, L., and Myllyla, R. (2002) Identification of amino acids important for the catalytic activity of the collagen glucosyltransferase associated with the multifunctional lysyl hydroxylase 3 (LH3). *J Biol Chem* 277, 18568-18573
107. Ruotsalainen, H., Sipila, L., Vapola, M., Sormunen, R., Salo, A. M., Uitto, L., Mercer, D. K., Robins, S. P., Risteli, M., Aszodi, A., Fassler, R., and Myllyla, R. (2006) Glycosylation catalyzed by lysyl hydroxylase 3 is essential for basement membranes. *J Cell Sci* 119, 625-635
108. Sipila, L., Ruotsalainen, H., Sormunen, R., Baker, N. L., Lamande, S. R., Vapola, M., Wang, C., Sado, Y., Aszodi, A., and Myllyla, R. (2007) Secretion and assembly of type IV and VI collagens depend on glycosylation of hydroxylysines. *J Biol Chem* 282, 33381-33388
109. Engvall, E., Hessle, H., and Klier, G. (1986) Molecular assembly, secretion, and matrix deposition of type VI collagen. *J Cell Biol* 102, 703-710
110. Maki, J. M., Tikkanen, H., and Kivirikko, K. I. (2001) Cloning and characterization of a fifth human lysyl oxidase isoenzyme: the third member of the lysyl oxidase-related subfamily with four scavenger receptor cysteine-rich domains. *Matrix Biology* 20, 493-496
111. Knott, L., and Bailey, A. J. (1998) Collagen cross-links in mineralizing tissues: a review of their chemistry, function, and clinical relevance. *Bone* 22, 181-187
112. Huber, M. A. (2007) Osteogenesis imperfecta. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 103, 314-320
113. Danlos, H. (1908) Un cas de Cutis laxa avec tumeurs par contusion chronique des coudes et des genoux (xanthome juvénile pseudo-diabétique de MM. Hallopeau et Macé de Lépinay). *Bull Soc Fr Dermatol Syphilig* 19, 70-72
114. Beighton, P., De Paepe, A., Steinmann, B., Tsipouras, P., and Wenstrup, R. J. (1998) Ehlers-Danlos syndromes: revised nosology, Villefranche, 1997. Ehlers-Danlos National Foundation (USA) and Ehlers-Danlos Support Group (UK). *Am J Med Genet* 77, 31-37
115. Lawrence, E. J. (2005) The clinical presentation of Ehlers-Danlos syndrome. *Adv Neonatal Care* 5, 301-314

116. Glorieux, F. H., Ward, L. M., Rauch, F., Lalic, L., Roughley, P. J., and Travers, R. (2002) Osteogenesis imperfecta type VI: a form of brittle bone disease with a mineralization defect. *J Bone Miner Res* 17, 30-38
117. Glorieux, F. H., Rauch, F., Plotkin, H., Ward, L., Travers, R., Roughley, P., Lalic, L., Glorieux, D. F., Fassier, F., and Bishop, N. J. (2000) Type V osteogenesis imperfecta: a new form of brittle bone disease. *J Bone Miner Res* 15, 1650-1658
118. Ward, L. M., Rauch, F., Travers, R., Chabot, G., Azouz, E. M., Lalic, L., Roughley, P. J., and Glorieux, F. H. (2002) Osteogenesis imperfecta type VII: an autosomal recessive form of brittle bone disease. *Bone* 31, 12-18
119. Plotkin, H. (2004) Syndromes with congenital brittle bones. *BMC Pediatr* 4, 16
120. Millington-Ward, S., McMahon, H. P., and Farrar, G. J. (2005) Emerging therapeutic approaches for osteogenesis imperfecta. *Trends Mol Med* 11, 299-305
121. Bank, R. A., Robins, S. P., Wijmenga, C., Breslau-Siderius, L. J., Bardoel, A. F. J., Van der Sluijs, H. A., Pruijs, H. E. H., and TeKoppele, J. M. (1999) Defective collagen crosslinking in bone, but not in ligament or cartilage, in Bruck syndrome: Indications for a bone-specific telopeptide lysyl hydroxylase on chromosome 17. *PNAS* 96, 1054-1058
122. Ha-Vinh, R., Alanay, Y., Bank, R. A., Campos-Xavier, A. B., Zankl, A., Superti-Furga, A., and Bonafe, L. (2004) Phenotypic and molecular characterization of Bruck syndrome (osteogenesis imperfecta with contractures of the large joints) caused by a recessive mutation in PLOD2. *Am J Med Genet A* 131, 115-120
123. Chatterjee, I. B., Kar, N. C., Ghosh, N. C., and Guha, B. C. (1961) Biosynthesis of L-ascorbic acid: missing steps in animals incapable of synthesizing the vitamin. *Nature* 192, 163-164
124. Olmedo, J. M., Yiannias, J. A., Windgassen, E. B., and Gornet, M. K. (2006) Scurvy: a disease almost forgotten. *Int J Dermatol* 45, 909-913
125. Peterkofsky, B. (1991) Ascorbate requirement for hydroxylation and secretion of procollagen: relationship to inhibition of collagen synthesis in scurvy. *Am J Clin Nutr* 54, 1135S-1140S
126. Robert, A. M., Godeau, G., Miskulin, M., and Moati, F. (1977) Mechanism of action of collagenase on the permeability of the blood-brain barrier. *Neurochem Res* 2, 449-455
127. Tilling, T., Korte, D., Hoheisel, D., and Galla, H. J. (1998) Basement membrane proteins influence brain capillary endothelial barrier function in vitro. *J Neurochem* 71, 1151-1157
128. Urabe, N., Naito, I., Saito, K., Yonezawa, T., Sado, Y., Yoshioka, H., Kusachi, S., Tsuji, T., Ohtsuka, A., Taguchi, T., Murakami, T., and Ninomiya, Y. (2002) Basement membrane type IV collagen molecules in the choroid plexus, pia mater and capillaries in the mouse brain. *Arch Histol Cytol* 65, 133-143
129. Seppanen, A., Suuronen, T., Hofmann, S. C., Majamaa, K., and Alafuzoff, I. (2007) Distribution of collagen XVII in the human brain. *Brain Res* 1158, 50-56

130. Deprez, P., and Inestrosa, N. C. (2000) Molecular modeling of the collagen-like tail of asymmetric acetylcholinesterase. *Protein Eng* 13, 27-34
131. Anglister, L., and Silman, I. (1978) Molecular structure of elongated forms of electric eel acetylcholinesterase. *J Mol Biol* 125, 293-311
132. Lwebuga-Mukasa, J. S., Lappi, S., and Taylor, P. (1976) Molecular forms of acetylcholinesterase from Torpedo californica: their relationship to synaptic membranes. *Biochemistry* 15, 1425-1434
133. Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., and Claverie, J. M. (2004) The 1.2-megabase genome sequence of Mimivirus. *Science* 306, 1344-1350
134. Xiao, C., Chipman, P. R., Battisti, A. J., Bowman, V. D., Renesto, P., Raoult, D., and Rossmann, M. G. (2005) Cryo-electron microscopy of the giant Mimivirus. *J Mol Biol* 353, 493-496
135. Patenaude, S. I., Seto, N. O., Borisova, S. N., Szpacenko, A., Marcus, S. L., Palcic, M. M., and Evans, S. V. (2002) The structural basis for specificity in human ABO(H) blood group biosynthesis. *Nat Struct Biol* 9, 685-690

6 Acknowledgement

First of all, I would like to thank my thesis supervisor Prof. Dr. Thierry Hennet for giving me the opportunity to perform my PhD thesis on this magnificent project and for his support during the last years.
Furthermore, I would like to thank my thesis committee, Prof. Dr. Markus Aebi and Prof. Dr. Sabine Werner for their interest in my project and the helpful discussions.
I am thankful to Prof. Dr. med. Beat Steinmann and Dr. Cecilia Giunta for providing the patient fibroblasts and for the interesting discussions about collagen.
Many thanks go to all researchers from the L-floor for the excellent working atmosphere. Furthermore, I would like to thank all members of the Hennet group, especially Charlotte Maag and Andrea Fuhrer, two really great friends in good and bad times, and Christoph Rutschmann, the only man in our lab. He was really a good support to me and a very good help for the project even though it was not always easy to work with the collagen. Stefan Deuber, who well established the MIMIV project which I took over from him after his departure. Also many thanks to Albana Rexhepaj for her friendship, her interest and her support and to Andreas Hülsmeier who did the MS and introduced me to the HPLC. Many thanks go also to the Berger, Borsig and Rohrer groups which were also always open for questions and giving useful advice.
I am thankful to Claudia Zeissig, with whom I had many good conversations during an infinite number of ergotherapy after having a hand-cast for 11 weeks. She helped me really a lot. I would not have been able to continue my labwork without her very good therapy during 16 months. In this context I also want to say a big thank to Heinz Läubli who always encouraged and informed me with his knowledge about the scaphoid.
Many thanks go to my friends. I could always count on their friendship and talk about the hard time in the lab of a graduate student.
A special thank you goes to my family, who was always there for me, had an open ear and supported me in my education. I am deeply grateful to Jonas, for his sympathy, his help and his great love.

Curriculum Vitae

Name	Belinda Schegg
Date of birth	17th of December 1979
Place of birth	St.Gallen
Citizen of	Oberriet-Montlingen SG

Education

2004-2008	PhD thesis under the guidance of Prof. Dr. Thierry Hennet at the Institute of Physiology, University of Zurich and the thesis supervisor Prof. Dr. Markus Aebi, Institute of Microbiology, Swiss Federal Institute of Technology (ETH), Zurich
1999-2004	Studies of biology at the Swiss Federal Institute of Technology (ETH), Zurich Diploma Thesis at the Institute of Microbiology, Prof. Markus Aebi
1995-1999	Gymnasium in St.Gallen Final examination: Matura Typus B
1993-1995	Secondary Education in St.Gallen
1987-1993	Primary Education in St.Gallen

Publications

2005	**Glycobiology 2005 Apr;15(4):361-7** The N-X-S/T consensus sequence is required but not sufficient for bacterial N-linked protein glycosylation. Nita-Lazar M, Wacker M, **Schegg B**, Amber S, Aebi M